



## DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITE DE COOPERATION EN MATIERE DE BREVETS (PCT)

(51) Classification internationale des brevets <sup>6</sup> : <b>C12N 15/82, 15/12, C07K 14/805, C12N 5/10, A01H 5/00, A61K 38/42</b>		A2	(11) Numéro de publication internationale: <b>WO 97/04115</b>  (43) Date de publication internationale: <b>6 février 1997 (06.02.97)</b>
(21) Numéro de la demande internationale: <b>PCT/FR96/01123</b>		F-63870 Orcines (FR). POYART, Claude [FR/FR]; 67, rue Croulebarbe, F-75013 Paris (FR).	
(22) Date de dépôt international: <b>17 juillet 1996 (17.07.96)</b>		(74) Mandataires: <b>GUTMANN, Ernest etc.; Ernest Gutmann-Yves Plasseraud S.A., 3, rue Chauveau-Lagarde, F-75008 Paris (FR).</b>	
(30) Données relatives à la priorité: <b>95/08675 17 juillet 1995 (17.07.95) FR</b>		(81) Etats désignés: <b>AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, brevet ARIPO (KE, LS, MW, SD, SZ, UG), brevet eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</b>	
(71) Déposants ( <i>pour tous les Etats désignés sauf US</i> ): <b>BIOCEM [FR/FR]; Campus Universitaire des Cézeaux, 24, avenue des Landais, F-63170 Aubière (FR). INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE MEDICALE [FR/FR]; 101, rue de Tolbiac, F-75654 Paris Cédex 13 (FR).</b>		Publiée <i>Sans rapport de recherche internationale, sera republiée dès réception de ce rapport.</i>	
(72) Inventeurs; et (75) Inventeurs/Déposants ( <i>US seulement</i> ): <b>MEROT, Bertrand [FR/FR]; 3, la Coussedière, F-63530 Volvic (FR). DIERYCK, Wilfrid [FR/FR]; 14, rue du Rhin, F-77178 Saint-Pathus (FR). LENEE, Philippe [FR/FR]; 4, rue Beaudelaire - Haut-Magenta, Boîte postale 14555, F-98800 Nouméa (FR). MARDEN, Michael [US/FR]; 21, avenue Gambetta, F-93600 Aulnay-sous-Bois (FR). GRUBER, Véronique [FR/FR]; 44, avenue Jean-Jaurès, F-63400 Chamalières (FR). PAGNIER, Renée-Josée [FR/FR]; 8, rue Jean-Monnet, F-94270 Le Kremlin-Bicêtre (FR). BAUDINO, Sylvie [FR/FR]; 2, lotissement des Volcans,</b>			

(54) Title: **METHOD FOR PRODUCING HAEMIN PROTEINS USING PLANT CELLS, RESULTING PROTEINS AND PRODUCTS CONTAINING SAME**

(54) Titre: **PROCEDE DE PRODUCTION, PAR DES CELLULES VEGETALES, DE PROTEINES HEMINIQUES, PROTEINES AINSI OBTENUES ET PRODUITS CONTENANT CES PROTEINES**

(57) Abstract

A method for producing haemin proteins by (i) inserting into plant cells one or more nucleic acid molecules that each comprise at least one sequence coding for a protein component of an animal haemin protein capable of reversibly binding oxygen, or for a variant or portion of said protein component, and optionally a sequence coding for a selection agent; (ii) selecting cells containing nucleic acid coding for the protein component of the haemin protein; (iii) optionally propagating the transformed cells either in a culture or by regenerating whole transgenic or chimeric plants; and (iv) recovering and optionally purifying a haemin protein that includes a complex consisting of the protein or proteins coded for by said nucleic acid and at least one iron-containing porphyritic nucleus, or a plurality of such complexes.

(57) Abrégé

L'invention concerne un procédé de production de protéines héminiques comprenant les étapes suivantes: i) introduction, dans des cellules végétales, d'une ou plusieurs molécule(s) d'acides nucléiques dont chacune comporte au moins une séquence codant pour un composant protéique d'une protéine héminique d'origine animale capable de fixer réversiblement l'oxygène, ou pour une variante ou une partie de ce composant protéique, et éventuellement une séquence codant pour un agent de sélection; ii) sélection des cellules contenant l'acide nucléique codant pour le composant protéique de la protéine héminique; iii) éventuellement, propagation des cellules transformées, soit en culture, soit par la régénération de plantes entières transgéniques ou chimériques; iv) récupération, et éventuellement purification, d'une protéine héminique comprenant un complexe constitué de la protéine ou des protéines codées par le susdit acide nucléique et au moins un noyau porphyrique à fer, ou une pluralité de ces complexes.

**UNIQUEMENT A TITRE D'INFORMATION**

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publient des demandes internationales en vertu du PCT.

AT	Arménie	GB	Royaume-Uni	MW	Malawi
AT	Autriche	GE	Géorgie	MX	Mexique
AU	Australie	GN	Guinée	NE	Niger
BB	Barbade	GR	Grèce	NL	Pays-Bas
BE	Belgique	HU	Hongrie	NO	Norvège
BF	Burkina Faso	IE	Irlande	NZ	Nouvelle-Zélande
BG	Bulgarie	IT	Italie	PL	Pologne
BJ	Bénin	JP	Japon	PT	Portugal
BR	Brésil	KE	Kenya	RO	Roumanie
BY	Bélarus	KG	Kirghizistan	RU	Fédération de Russie
CA	Canada	KP	République populaire démocratique de Corée	SD	Soudan
CF	République centrafricaine	KR	République de Corée	SE	Suède
CG	Congo	KZ	Kazakhstan	SG	Singapour
CH	Suisse	LI	Liechtenstein	SI	Slovénie
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovaquie
CM	Cameroun	LR	Libéria	SN	Sénégal
CN	Chine	LT	Lituanie	SZ	Swaziland
CS	Tchécoslovaquie	LU	Luxembourg	TD	Tchad
CZ	République tchèque	LV	Lettanie	TG	Togo
DE	Allemagne	MC	Monaco	TJ	Tadjikistan
DK	Danemark	MD	République de Moldova	TT	Trinité-et-Tobago
EE	Estonie	MG	Madagascar	UA	Ukraine
ES	Espagne	ML	Mali	UG	Ouganda
FI	Finlande	MN	Mongolie	US	Etats-Unis d'Amérique
FR	France	MR	Mauritanie	UZ	Ouzbékistan
GA	Gabon			VN	Viet Nam

**PROCEDE DE PRODUCTION, PAR DES CELLULES VEGETALES, DE  
PROTEINES HEMINIQUES, PROTEINES AINSI OBTENUES ET  
PRODUITS CONTENANT CES PROTEINES**

L'invention concerne un procédé de production, par des cellules végétales, de protéines héminiques, et en particulier les protéines héminiques capables de fixer réversiblement l'oxygène, par exemple l'hémoglobine et ses dérivés, et la myoglobine. Elle concerne en outre les protéines obtenues par la mise en oeuvre du procédé. L'invention vise également les cellules et plantes génétiquement transformées capable de produire ces protéines, et les construits d'acides nucléiques impliqués dans la transformation. En outre, l'invention vise des produits par exemple, pharmaceutiques ou cosmétiques, contenant ces protéines héminiques.

Les protéines héminiques sont des molécules complexes composées d'une ou plusieurs chaîne(s) polypeptidique(s) en association à un ou plusieurs noyau(x) ferroporphyrrique(s). Ces noyaux sont composés de quatre cycles pyrrole, juxtaposés en une structure fermée et reliés par des ponts méthéniques, et contenant au centre de la molécule, un atome de fer. Les protéines héminiques se distinguent les unes des autres par la nature et le nombre des chaînes polypeptidiques et par la nature des chaînes latérales portées par les huit atomes  $\beta$  des cycles pyrrole. Un exemple de noyau ferroporphyrrique est la protoporphyrine IX à fer également connue sous le nom de "protohème" ou simplement "hème" (figure 1).

La famille des protéines héminiques comprend de nombreuses substances importantes sur le plan

biologique chez les animaux et chez les plantes, notamment l'hémoglobin, la myoglobine, les cytochromes, les peroxydases, les catalases.

L'hémoglobin est le constituant principal des globules rouges du sang. Elle a pour fonction essentielle de fixer, transporter et délivrer la quantité d'oxygène nécessaire au fonctionnement normal des tissus.

La molécule d'hémoglobin est composée de deux types de chaînes ou sous-unités de globine, nommées  $\alpha$  et  $\beta$  (de 141 et 146 acides aminés respectivement), et appariées deux à deux pour former un tétramère  $\alpha_2\beta_2$ . Chacune de ces sous-unités renferme, solidement attachée dans une poche hydrophobe, une molécule d'hème (c'est-à-dire la protoporphyrine IX) contenant en son centre un atome de fer divalent ( $Fe^{2+}$ ) auquel se lie de façon réversible une molécule d'oxygène. Chaque molécule d'hémoglobin tétramérique contient donc 4 atomes de fer et 4 molécules d'oxygène qu'elle fixe lors de son passage dans les poumons. La masse moléculaire du tétramère est de 64 650 D. Chez l'homme, les chaînes  $\alpha$  et  $\beta$  sont synthétisées à partir de deux types de gènes situés sur les chromosomes 16 et 11 respectivement.

Le terme de chaînes de type bêta, ou "non alpha", recouvre, non seulement les chaînes bêta, mais également les chaînes dénommées epsilon, gamma ou delta.

Normalement, chez l'adulte, plus de 95% de l'hémoglobin est constituée par du tétramère  $\alpha_2\beta_2$ , c'est-à-dire l'association de deux dimères hétérologues alpha-bêta, associés au complexe catalytique, l'hème. On trouve 2% à 3% d'une hémoglobin constituée de tétramères  $\alpha_2\delta_2$ , et des traces d'hémoglobin fo tale  $\alpha_2\gamma_2$ .

La molécule d'hémoglobine humaine tétramérique existe sous deux formes ou structures quaternaires selon que l'oxygène est lié ou non aux atomes de fer. En présence d'oxygène, l'hémoglobine est dite en état R (pour relâchée) et son affinité pour l'oxygène est forte. En l'absence d'oxygène, l'hémoglobine est dite en état T (pour tendue) et son affinité pour l'oxygène est 100 fois plus faible (Perutz, 1970). L'affinité résultante est liée à l'équilibre entre les concentrations de formes R et T. Cette affinité est d'autant plus faible que la concentration de l'hémoglobine sous forme T est plus élevée à tout niveau d'oxygénéation. L'affinité de l'hémoglobine pour l'oxygène est réglée par le co-facteur - 2,3-diphosphoglycérate (DPG), une petite molécule provenant du métabolisme du glucose et qui se fixe sur les chaînes  $\beta$  de l'hémoglobine tétramérique diminuant son affinité pour l'oxygène.

L'augmentation des risques infectieux par des produits dérivés du sang humain (hépatites, H.I.V) rend nécessaire la mise au point d'un transporteur artificiel d'oxygène comme substitut à la transfusion sanguine.

Des techniques mettant en oeuvre les ADN recombinants ont été proposées pour produire les chaînes protéiques de la globine.

Les premières techniques mises en oeuvre avaient essentiellement pour but de faire synthétiser dans E. coli, les chaînes alpha et bêta séparément (Nagaï et Thogen-Sen, 1987), impliquant des procédés lourds d'expression séparée de chacune des chaînes. Ces procédés pouvaient difficilement être exploités à l'échelle industrielle.

Plus récemment, l'expression d'hémoglobine recombinante soluble et fonctionnelle a été mise au

point dans E. coli (Hoffman et al., 1990, P.N.A.S., 87, 8521-8525) et Saccharomyces cerevisiae (Wagenbach et al., 1991, Biotechnology, 9, 57-61). Chacun de ces systèmes présente avantages et inconvénients. En effet, les taux d'expression les plus élevés sont obtenus dans E. coli qui présente néanmoins le désavantage de produire des endotoxines et de ne pas cliver les méthionines NH<sub>2</sub>-terminales contrairement à Saccharomyces cerevisiae. Dans la levure, les rendements de synthèse d'hémoglobine sont faibles (3 à 5%), comparés aux 10-15% obtenus dans E. coli. Ceci limite actuellement l'utilisation de la levure dans le cadre d'un projet de développement industriel.

L'utilisation de cellules animales en culture ou d'animaux transgéniques comme hôtes de production a également été réalisée (Swanson et al., Bio/Technology, Mai 1992, 10, page 55). Ces techniques ne paraissent pas exploitables actuellement en raison des faibles taux d'expression et les risques de contaminations virales et par des prions.

Le problème technique que se propose de résoudre la présente invention est de produire des protéines héminiques, et notamment l'hémoglobine et ses dérivés, en grande quantité à des coûts faibles, sans risque de contaminations virales ou sub-virales. Les inventeurs ont apporté une solution à ce problème en utilisant comme hôte pour la transformation et la production, des cellules végétales.

Diverses équipes se sont déjà intéressées à la production de protéines recombinantes de mammifères dans les cellules végétales ou dans des plantes transgéniques. Par exemple, l'expression spécifique dans les graines de colza, de la leu-enképhaline a été obtenue avec des niveaux d'expression d'environ 0,1%

(Vanderkerckhove et al., Biotechnology, 1989, 7, 929-932).

En 1990, Sijmons et al., (Biotechnology, 1990, 8, 217-221) ont transféré le gène de la sérum albumine humaine dans des cellules de tabac et de pomme de terre. Quelle que soit l'origine des peptides signaux (humaine ou végétale), des taux de sérum albumine humaine de l'ordre 0,02% des protéines totales ont été obtenus dans les feuilles, les tiges et les tubercules de pomme de terre.

D'autres protéines recombinantes de mammifères ont été aussi produites dans les plantes : l'antigène de surface de l'hépatite B (Mason et al., P.N.A.S., 1992, 89, 11745-11749) ; l'interféron humain (Edelbaum J. of Interferon Res., 1992, 12, 449-453) ; un anticorps de souris anti-Streptococcus mutans, agent de la carie dentaire (Hiatt et Ma, FEBS, 1992, 307, 71-75) ; un anticorps anti-Herpès (Russel, 1994) et l'hirudine (Moloney, 1994).

L'ensemble de ces recherches montre que la production de protéines recombinantes de mammifères dans les cellules végétales est possible et que les mécanismes de synthèse de protéines à partir des séquences d'ADN sont similaires chez les cellules animales et les cellules végétales.

En revanche, il existe peu d'informations au sujet des porphyrines à fer chez les végétaux, notamment sur leurs structures, leurs voies de synthèse et l'assemblage des noyaux porphyriques et les chaînes protéiques pour former les protéines héminiques. La production de molécules recombinantes ayant la capacité de fixer réversiblement l'oxygène, et nécessitant l'assemblage, dans la cellule, de protéines hétérologues et de porphyrines endogènes végétales n'a jamais été décrite.

L'invention concerne un procédé de production, par des cellules végétales, de protéines héminiques recombinantes. Selon le procédé de l'invention, la cellule végétale est modifiée génétiquement afin de pouvoir exprimer le composant protéique d'une protéine héminique. Le noyau porphyrique est produit par la cellule de manière endogène, l'assemblage des composants protéiques et porphyriques ayant lieu spontanément grâce à leur affinité élevée l'un pour l'autre.

Plus particulièrement, l'invention concerne un procédé de production de protéines héminiques comprenant les étapes suivantes :

- i) introduction, dans des cellules végétales, d'une ou plusieurs molécule(s) d'acide nucléique dont chacune comporte au moins une séquence codant un composant protéique d'une protéine héminique d'origine animale ou pour une variante ou une partie de ce composant protéique, et éventuellement une séquence codant un agent de sélection ;
- ii) sélection des cellules ayant intégré l'acide nucléique codant pour le composant protéique ;
- iii) propagation des cellules transformées, soit en culture, soit par la régénération de plantes entières transgéniques ou chimériques ;
- iv) récupération, et éventuellement purification, d'une protéine héminique comprenant un assemblage de la protéine ou des protéines codées par le susdit acide nucléique avec au moins un noyau porphyrique à fer, ou une pluralité de ces assemblages.

L'invention concerne de préférence un procédé de production de protéines héminiques comprenant les étapes suivantes :

- i) introduction, dans des cellules végétales, d'une ou plusieurs molécule(s) d'acides nucléiques dont chacune

comporte au moins une séquence codant pour un composant protéique d'une protéine héminique d'origine animale de préférence capable de fixer réversiblement l'oxygène, ou pour une variante ou une partie de ce composant protéique, et éventuellement une séquence codant pour un agent de sélection ;  
ii) sélection des cellules contenant l'acide nucléique codant pour le composant protéique de la protéine héminique ;  
iii) éventuellement, propagation des cellules transformées, soit en culture, soit par la régénération de plantes entières transgéniques ou chimériques ;  
iv) récupération, et éventuellement purification, d'une protéine héminique comprenant un complexe constitué de la protéine ou des protéines codées par le susdit acide nucléique et au moins un noyau porphyrique à fer, ou une pluralité de ces complexes.

Dans le contexte de la présente invention, le terme "protéine héminique" signifie toute protéine ayant un noyau porphyrique à fer comme groupement prosthétique, et en particulier la protoporphyrine IX telle qu'elle existe dans l'hémoglobine et la myoglobine humaine (figure 1). Le noyau porphyrique peut également être des dérivés de l'hème de celles de l'hème humaine. De préférence, les chaînes latérales sont hydrophobes.

Les protéines héminiques de l'invention englobent en particulier les protéines héminiques ayant comme fonction principale la fixation réversible de l'oxygène, c'est-à-dire la myoglobine et l'hémoglobine, ainsi que les cytochromes dont le rôle est de transporter des électrons. Les dérivés de ces protéines conservant ces fonctionnalités sont aussi englobés par l'invention.

Selon une variant préférée, la protéine héminique de l'invention est l'hémoglobine ou une protéine de type hémoglobine. Dans le contexte de l'invention, le terme "protéine de type hémoglobine" englobe toutes les protéines héminiques présentant à la fois :

- i) une ou plusieurs chaîne(s) d' $\alpha$ - et/ou de  $\beta$ -globine, ou de variantes de ces polypeptides, et
- ii) une ou plusieurs molécules de protoporphyrine IX à fer, ou de protoporphyrines se distinguant de la protoporphyrine IX par la nature des chaînes latérales,
- iii) ayant une capacité de fixer réversiblement l'oxygène, de préférence avec une affinité comprise entre 10 et 50 mm Hg à 37°C, pH7.4. Plus particulièrement, l'affinité est comprise entre 20 et 30 mm Hg, à titre d'exemple, la  $P_{50}$  du sang total à pH 7.2 et de  $26 \pm 2$  mm Hg.

Dans ce qui suit, le terme "molécule de type hémoglobine" sera utilisé de manière synonyme avec le terme "dérivé de l'hémoglobine".

Dans ce contexte, une "variante" d'un composant protéique, et particulièrement de l' $\alpha$ - ou  $\beta$ -globine, signifie une séquence en acides aminés qui se différencie par rapport à la séquence naturelle, par une ou plusieurs substitution(s), délétion(s) ou insertion(s) d'acides aminés. En général, la variante présente au moins 90%, et de préférence au moins 95% d'homologie, ou d'identité avec la séquence naturelle. Dans le contexte de la présente invention, le pourcentage d'homologie entre deux séquences d'acides aminés est calculé comme étant le nombre d'acides aminés identiques plus le nombre d'acides aminés similaires dans l'alignement des deux séquences, divisé par la longueur des séquences entre deux

positions données. Si, entre les deux positions données, les deux séquences n'ont pas la même longueur, le pourcentage d'homologie est le nombre d'acides aminés identiques et similaires, divisé par la longueur de la séquence la plus longue. Les acides aminés considérés comme étant "similaires" sont connus dans l'art, voir par exemple R.F. Feng, M.S. Jobson and R.F. Doolittle ; J. Mol. Evol. ; 1985 ; 21 ; 112-115. Ils sont normalement considérés comme étant ceux qui, au sein d'une matrice de permutation, ont un coefficient de substitution positif.

Le terme "variante" englobe aussi des fragments de chaînes polypeptidiques, par exemple de l' $\alpha$ - ou de la  $\beta$ -globine, ayant normalement une longueur d'au moins 90% de la molécule mère. Les variantes peuvent également être rendues plus longues que la molécule mère par l'addition de séquences non-fonctionnelles. De préférence, les variantes conservent les propriétés biologiques et immunologiques de la molécule mère.

La première étape du procédé de l'invention consiste en l'introduction, dans des cellules végétales, d'une ou plusieurs molécule(s) d'acide nucléique comportant au moins une séquence codant un composant protéique d'une protéine héminique de mammifère, ou pour une variante de ce composant.

Lorsque la protéine héminique est une protéine à chaîne unique, par exemple la myoglobine ou le cytochrome, l'acide nucléique introduit dans les cellules végétales comporte normalement une copie de la séquence codant pour cette protéine.

En revanche, lorsqu'il s'agit d'une protéine oligomérique ou multimérique, telle que l'hémoglobine ou des molécules du type hémoglobine, les séquences codant les différentes unités protéiques sont introduites dans la cellule végétale, soit au sein de

la même molécule d'acide nucléique, soit au sein de molécules d'acides nucléiques distinctes. De préférence, pour la production d'hémoglobine et ses dérivés, les séquences codant l' $\alpha$ - et  $\beta$ -globine, ou leurs variantes, se trouvent au sein du même vecteur, dit vecteur de co-expression. Le vecteur peut comporter une ou plusieurs copie(s) de chaque séquence codante.

Alternativement, les séquences codant l' $\alpha$ - et la  $\beta$ -globine, ou leurs variantes, peuvent se trouver sur des molécules d'acides nucléiques distinctes. Selon cette variante, les deux molécules peuvent être introduites dans la même cellule végétale, à condition de disposer d'un système de sélection approprié. Une autre technique consiste en l'introduction de l'une des molécules dans une première cellule végétale, et l'autre dans une deuxième cellule végétale. Chacune des cellules transformées est ensuite régénérée en plante entière, les plantes ainsi obtenues pouvant alors être croisées pour donner lieu à une descendance capable de produire à la fois les chaînes  $\alpha$  et  $\beta$ . Cette approche peut être utilisée pour optimiser le rendement de l'hémoglobine.

Les molécules d'acide nucléique introduites dans la cellule végétale lors de la première étape du procédé font également partie de l'invention. De manière générale, ces acides nucléiques comportent :

- i) une ou plusieurs séquence(s) codant un composant protéique d'une protéine héminique animale, et
- ii) une ou plusieurs séquence(s) codant un signal d'adressage d'origine végétale, et/ou des séquences régulatrices de transcription reconnues par une cellule végétale.

Plus particulièrement, l'acide nucléique de l'invention comporte :

- i) une ou plusieurs séquence(s) codant un composant protéique d'une protéine héminique animale, ladite protéine ayant la capacité de fixer réversiblement l'oxygène, et
- ii) des séquences régulatrices de transcription reconnues par une cellule végétale, comprenant un promoteur et des séquences régulatrices de terminaison, et
- iii) une ou plusieurs séquence(s) codant un signal d'adressage d'origine végétale.

De préférence, les séquences codant le composant protéique codent l' $\alpha$ - ou la  $\beta$ -globine animale, par exemple humaine ou bovine, ou les variantes de celles-ci. De cette façon, les propriétés de la molécule, et notamment l'affinité pour l'oxygène et la stabilité, peuvent être optimisées.

Parmi ces modifications, il est possible par exemple d'introduire dans l'une ou dans les deux chaînes  $\alpha$ - et  $\beta$ -globine, par mutagénèse dirigée, une ou deux différence(s) de séquences pour en diminuer l'affinité pour l'oxygène. Ces mutations peuvent être choisies à partir d'exemples de mutations naturelles (voir tableau I), soit des mutations indiquées par l'examen du modèle tridimensionnel de l'hémoglobine A naturelle.

Tableau I : Quelques hémoglobines humaines mutées  
 (Int. Hemoglobin Center, 1995)

Hémoglobine anormale	Positions et résidus normaux	Remplacement
chaîne α		
I	16 Lys	Glu
G <sub>Honolulu</sub>	30 Glu	Gln
Norfolk	57 Gly	Asp
M <sub>Boston</sub>	58 His	Tyr
G <sub>Philadelphia</sub>	68 Asn	Lys
O <sub>Indonesia</sub>	116 Glu	Lys
chaîne β		
C	6 Glu	Lys
S	6 Glu	Val
G <sub>San José</sub>	7 Glu	Gly
E	26 Glu	Lys
M <sub>Saskatoon</sub>	63 His	Tyr
Zürich	63 His	Arg
M <sub>Milwaukee</sub>	67 Val	Glu
D <sub>Punjab</sub>	121 Glu	Gln
Mequon	41 Phe	Tyr
Providence	82 Lys	Asp

D façon très avantageuse, on utilisera les mutants dont les propriétés fonctionnelles correspondent aux conditions physiologiques du transport de l'oxygène : liaison réversible, coopérativité et faible vitesse d'autooxydation. Parmi les mutants, on utilisera de préférence les doubles mutants  $\alpha_2\beta_2F41Y, K82D$  (c'est-à-dire un mutant dont la chaîne  $\beta$  comporte les modifications suivantes : Phe-41 est remplacé par Tyr, et Lys-82 est remplacé par Asp) ou  $\alpha_2\beta_2F41Y, K66T$  (c'est-à-dire un mutant dont la chaîne  $\beta$  comporte les modifications suivantes : Phe-41 est remplacé par Tyr, et Lys-66 est remplacé par Thr) qui correspondent à ces caractéristiques fonctionnelles.

La modification des chaînes  $\alpha$  et  $\beta$  peut aussi être effectuée afin de stabiliser la molécule, c'est-à-dire d'éviter la dissociation du tétramère en dimères de petites tailles rapidement filtrés par les reins et limitant la durée de vie intra-vasculaire de l'hémoglobine. Le pontage covalent, à l'aide de phosphate ou de diaspirine, a été démontré comme étant une technique efficace pour stabiliser le tétramère (Benesch et Kwong, 1994). Le même résultat peut être obtenu par le biais de modifications de la chaîne en acides aminés. Les sous-unités  $\alpha$  sont produites sous forme dimérique alpha-alpha liée par un résidu glycyl. Sous cette forme, elles conservent leur capacité de s'assembler correctement aux sous-unités partenaires bêta et à l'hème pour former une hémoglobine soluble. Cette hémoglobine ne peut plus se dissocier en dimères car la structure tétramérique est stabilisée par une liaison covalente (liaison peptidique) entre les dimères alpha-bêta. Cette technique permet d'augmenter la demi-vie intra-vasculaire de la molécule.

Parmi les variantes, il est également possible d'utiliser une protéine hybride composée d'une partie de la chaîne alpha et une partie de la chaîne bêta.

Selon une variante préférée de l'invention, l'acide nucléique comprend, outre les séquences codant l' $\alpha$ - ou  $\beta$ -globine, des séquences codant des signaux d'adressage. De préférence, ces signaux sont des signaux d'adressage chloroplastiques ou mitochondriaux. L'expression et/ou l'accumulation des protéines recombinantes dans ces organites est particulièrement préférée en raison de la disponibilité de porphyrines à fer endogènes que l'on trouve ici. Le rendement en protéines héméniques est donc augmenté. En outre, l'adressage des protéines vers les chloroplastes et les mitochondries évite la glycosylation de la protéine, ce qui peut être avantageux dans la mesure où la molécule d'hémoglobine naturelle n'est pas glycosylée.

Comme exemple de signaux d'adressage chloroplastiques, on peut citer la séquence codant pour le peptide transit du précurseur de la petite sous-unité de la ribulose 1,5-biphosphate carboxylase de Pisum sativum (voir exemples). Comme signaux d'adressage mitochondrial, on peut citer la séquence codant pour le peptide transit du précurseur de la sous-unité bêta de l'ATP-aseF1 mitochondriale de Nicotiana plumbaginifolia (voir exemples).

Ces peptides transits, ainsi que la méthionine N-terminale, sont normalement clivés dans les chloroplastes ou les mitochondries. L'expression des protéines dans les plastides a donc également l'avantage de produire une molécule dépourvue de la méthionine N-terminale comme la molécule naturelle.

Selon une autre variante, les séquences d'adressage peuvent être des séquences codant un

peptide signal N-terminal ("prépeptide"), éventuellement en association à un signal responsable de la rétention de la protéine dans le réticulum endoplasmique (signal du type KDEL), ou un signal d'adressage vacuolaire ou "propeptide". La présence du peptide signal N-terminal ou prépeptide permet la pénétration de la protéine naissante dans le réticulum endoplasmique où ont lieu un certain nombre de maturations post-traductionnelles, notamment le clivage du peptide signal, les N-glycosylations, si la protéine en question présente des sites de N-glycosylation, et la formation des ponts disulfures. Parmi ces différents signaux, le prépeptide, responsable de l'adressage de la protéine dans le réticulum endoplasmique, est dominant. Il s'agit normalement d'un peptide signal N-terminal hydrophobe ayant entre 10 et 40 acides aminés et étant d'origine animale ou végétale. De préférence, il s'agit d'un prépeptide d'origine végétale, par exemple celui de la sporamine, de la lectine d'orge, de l'extensine végétale, de l' $\alpha$ -mating factor, de la protéine de pathogénèse 1 ou 2.

Normalement, le peptide signal est clivé par un signal peptidase dès l'introduction co-traductionnelle du polypeptide naissant dans la lumière du RER. La protéine mature ne comporte plus cette extension N-terminale.

Les séquences d'adressage peuvent, outre le prépeptide, aussi comporter un signal de rétention endoplasmique, consistant en les peptides KDEL, SEKDEL ou HEKDEL. Ces signaux se trouvent normalement à l'extrémité C-terminale de la protéine et subsistent sur la protéine mature. La présence de ce signal a tendance à augmenter les rendements en protéines recombinantes.

Les signaux d'adressage peuvent, outre le prépeptide, comporter aussi un signal d'adressage vacuolaire ou "propeptide". En présence d'un tel signal, après passage dans le RER, la protéine est adressée aux vacuoles des tissus aqueux, par exemple les feuilles, ainsi qu'aux corps protéiques des tissus de réserve, par exemple les graines, tubercules et racines. L'adressage de la protéine vers les corps protéiques de la graine est particulièrement intéressant en raison de la capacité de la graine à accumuler des protéines, jusqu'à 40% des protéines par rapport à la matière sèche, dans des organites cellulaires dérivés des vacuoles, appelés corps protéiques et en raison de la possibilité de stocker plusieurs années les graines contenant les protéines recombinantes à l'état déshydraté.

Comme propeptide, on peut utiliser un signal d'origine animale ou végétale, les signaux végétaux étant particulièrement préférés, par exemple la prosporamine. Le propeptide peut être N-terminal ("N-terminal targeting peptide" ou NTTP), ou C terminal (CTTP). Dans la mesure où les propeptides sont normalement clivés dès l'entrée de la protéine dans la vacuole, il n'est pas présent dans la protéine mature.

L'utilisation du peptide signal ou prépeptide, peut conduire à la glycosylation de la protéine. Normalement, la globine n'a pas de sites de N-glycosylation, mais ceux-ci peuvent être introduits par mutagénèse. Les chaînes  $\alpha$  et  $\beta$  peuvent aussi présenter des sites de O-glycosylation.

En l'absence de tout signal d'adressage, la protéine est exprimée dans le cytoplasme.

L'acide nucléique introduit dans la cellule végétale peut aussi comprendre des séquences régulatrices de transcription reconnues par la cellule

végétale. L'acide nucléique est alors un "gène chimérique". Les séquences régulatrices comprennent un ou plusieurs promoteur(s) d'origine végétale, virale ou provenant d'Agrobacterium tumefaciens. Il peut s'agir de promoteurs constitutifs, par exemple le 35S du CaMV, le double 35S, les promoteurs Nos ou OCS, ou des promoteurs spécifiques de certains tissus comme le grain ou spécifiques de certaines phases de développement de la plante. Comme promoteurs spécifiques de graines, on peut citer le promoteur du gène de la napin et de l'acyl carrier protéine (ACP) (EP-A-0255378), ainsi que les promoteurs des gènes AT2S d'Arabidopsis thaliana, c'est-à-dire les promoteurs PAT2S1, PAT2S2, PAT2S3 et PAT2S4 (Krebbers et al., Plant Physiol., 1988, vol. 87, pages 859-866). Il est particulièrement préféré d'utiliser le promoteur de la cruciférine ou de la phaséoline ou pGEA1 et pGEA6 d'Arabidopsis, promoteurs de gènes de type "Em, Early Methionine labelled protein" fortement exprimée au cours des phases de dessication de la graine.

Il peut être envisagé d'utiliser des "enhancers" pour améliorer l'efficacité d'expression. Lorsque la transformation a lieu directement dans les génomes chloroplastiques et mitochondriaux, des promoteurs de gènes spécifiques de ces compartiments peuvent être utilisés.

Les séquences régulatrices de transcription comprennent normalement des séquences de terminaison de transcription qui sont d'origine végétale ou virale, par exemple 35S, ou bactérienne (Agrobacterium).

Lorsque l'acide nucléique transformant ne comprend pas de séquences régulatrices, il est préférable d'ajouter sur chaque extrémité de l'acide

nucléique, une séquence d'ADN homologue aux séquences génomiques qui jouxtent un site d'insertion particulier dans le génome. Ceci permet l'intégration du construct par recombinaison homologue, à un endroit où des séquences régulatrices endogènes peuvent contrôler l'expression des séquences hétérologues.

Les acides nucléiques de l'invention peuvent également comprendre un ou plusieurs intron(s), de préférence d'origine végétale. Ces introns provenant d'un gène végétal sont introduits artificiellement afin d'augmenter l'efficacité de l'expression de la séquence hétérologue.

En effet, il a été démontré, particulièrement chez les monocotylédones que l'insertion d'un intron dans la partie 5' non traduite d'un gène, c'est-à-dire entre le site d'initiation de transcription et le site d'initiation de traduction, conduit à une amélioration de la stabilité du messager, et par conséquence, à une meilleure expression. Le ou les introns utilisé(s) de cette manière proviennent de préférence d'une monocotylédone telle que le maïs. Il s'agit de préférence, mais pas obligatoirement, du premier intron du gène.

La séquence d'acides nucléiques codant pour l' $\alpha$ -et la  $\beta$ -globine et ses variantes est normalement de l'ADNc. Des séquences appropriées sont illustrées dans les figures 2 et 3, toute séquence dégénérée peut aussi être utilisée ainsi que les séquences des variantes telles que définies ci-dessus.

L'introduction de la ou des molécule(s) d'acides nucléiques dans la cellule végétale peut s'effectuer de manière stable soit par transformation du génome nucléaire, soit par transformation du génome chloroplastique de la cellule végétale, soit par transformation du génome mitochondrial.

Pour la transformation du génome nucléaire, des techniques classiques peuvent être mises en oeuvre. Tous les moyens connus pour introduire de l'ADN étranger dans les cellules végétales peuvent être utilisés, par exemple Agrobacterium, électroporation, fusion de protoplastes, bombardement avec canon à particules, ou pénétration d'ADN dans des cellules comme le pollen, la microspore, la graine et l'embryon immature. Les vecteurs viraux tels que les Geminivirus ou les virus satellites peuvent également être utilisés comme moyens d'introduction. Agrobacterium tumefaciens et rhizogenes constituent le moyen préféré. Dans ce cas, la séquence de l'invention est introduite dans un vecteur approprié avec toutes les séquences régulatrices nécessaires telles que promoteurs, terminateurs, etc... ainsi que toute séquence nécessaire pour sélectionner les transformants ayant intégré les séquences hétérologues.

La transformation du génome nucléaire de la cellule végétale est souvent effectuée en utilisant les signaux d'adressage mentionnés ci-dessus et qui déterminent le compartiment cellulaire où se fera l'expression et/ou l'accumulation de la protéine.

Selon une autre variante de l'invention, l'introduction de l'acide nucléique dans la cellule végétale peut être effectuée par la transformation des génomes mitochondriaux ou chloroplastiques (voir par exemple Carrer et al., Mol. Gen. Genet., 1993, 241, 49-56).

Les techniques de transformation directe des chloroplastes ou mitochondries sont connues en soi et peuvent comprendre les étapes suivantes :

20

- i) introduction de l'ADN transformant par la technique biolistique (Svab et al., P.N.A.S., 1990, 87, 8526-8530) ;
- ii) intégration de l'ADN transformant par deux évènements de recombinaison homologue ;
- iii) élimination sélective des copies du génome sauvage au cours de divisions cellulaires répétées sur milieu sélectif.

Afin de permettre la recombinaison homologue de l'ADN transformant, deux fragments d'ADN homologues aux séquences génomiques, par exemple les gènes rbcL et ORF 512 sont ajoutés à chaque extrémité de l'ADN à insérer dans le génome.

La transformation directe des chloroplastes ou mitochondries présente l'avantage d'augmenter sensiblement le rendement en hémoglobine mais la méthionine N-terminale est retenue.

Selon une autre variante de l'invention, l'acide nucléique hétérologue peut être introduit dans la cellule végétale par l'intermédiaire d'un vecteur viral.

Le procédé de l'invention comprend une étape de détection des protéines héminiques et notamment de l'hémoglobine et de ses dérivés. Ceci permet de vérifier si la plante ou la cellule végétale est capable non seulement d'exprimer les protéines hétérologues, mais aussi de les assembler correctement avec le noyau porphyrrique. Pour l'hémoglobine dans un environnement complexe contenant d'autres chromophores ou des molécules diffusant la lumière, on utilisera avantageusement la détection par spectroscopie optique résolue dans le temps. Cette technique est décrite en détail dans les exemples. D'autres techniques de détection consistent en l'utilisation d'anticorps spécifiques pour les chaînes alpha ou bêta globine, ou

leurs variantes. Les techniques spectrométriques et immunologiques peuvent être utilisées en association l'une avec l'autre. La mise en oeuvre de ces techniques permet de sélectionner les plantes qui sont capables de produire l'hémoglobine et ses dérivés selon l'invention.

Le procédé de l'invention comprend en outre une étape de récupération ou d'extraction de l'hémoglobine ou de ses dérivés des tissus végétaux. L'extraction est normalement faite par broyage des tissus, par exemple feuilles ou grains, dans un tampon approprié, filtrage du broyat, précipitation des protéines dans le surnageant, centrifugation et reprise du culot dans un tampon approprié avec dialyse. Une étape de purification partielle peut aussi être effectuée à ce stade par chromatographie sur colonne d'échangeuse d'anions.

Le tétramère d'hémoglobine, ou de ses dérivés, est purifié par deux chromatographies successives sur résine échangeuse d'ions suivies par une étape de concentration et saturation du concentrat en monoxyde de carbone. Ces techniques sont décrites en détail dans les exemples.

Lorsque l'expression de l'hémoglobine et de ses dérivés a lieu sous le contrôle d'un promoteur constitutif, tel que le promoteur double 35S, un taux d'expression d'au moins 1% d'hémoglobine par rapport aux protéines totales peut être obtenu. Les protéines représentent environ 10% de la masse sèche de la feuille et une tonne de feuilles sèches de tabac est récoltée par hectare. Il est donc possible d'obtenir de l'ordre de 100 grammes d'hémoglobine par hectare de tabac cultivé en supposant purifier seulement 10% de l'hémoglobine produite.

Le procédé de l'invention permet donc la production d'hémoglobine à des coûts très faibles avec une capacité de production plus élevée que celle obtenue en utilisant des fermentateurs pour la culture de bactéries ou de levure.

Outre le procédé de transformation, l'invention englobe aussi des vecteurs comprenant un ou plusieurs acide(s) nucléique(s) ou gène(s) chimérique(s) définis ci-dessus. Comme exemple de vecteurs, on peut citer des vecteurs binaires ou des plasmides, des vecteurs viraux tels que les geminivirus ou les CaMV.

L'invention concerne également les cellules végétales transformées avec les séquences d'acides nucléiques de l'invention. De préférence, il s'agit de cellules végétales transformées capables de produire une ou plusieurs hémoglobine(s) ou dérivés de l'hémoglobine selon l'invention.

Il peut s'agir de cultures de cellules végétales in vitro, par exemple en milieu liquide. Différents modes de culture ("batch", "fed batch" ou en continu) pour ce type de cellules sont actuellement à l'étude. Les cultures en "batch" sont comparables à celles effectuées en erlenmeyer dans la mesure où le milieu n'est pas renouvelé, les cellules ne disposent ainsi que d'une quantité limitée d'éléments nutritifs. La culture en "fed batch" correspond quant à elle à une culture en "batch" avec une alimentation programmée en substrat. Pour une culture en continu, les cellules sont alimentées en permanence avec du milieu nutritif. Un volume égal du mélange biomasse-milieu est ôté afin de maintenir le volume du réacteur constant. Les quantités de biomasse végétale envisageables avec des cultures en bioréacteurs sont variables selon l'espèce végétale, le mode de culture et le type de bioréacteur. Dans certaines conditions, des densités

de biomasse d'environ 10 à 30 g de poids sec par litre de cultur peuvent être obtenus, pour des espèces comme Nicotiana tabacum, Vinca rosea et Catharanthus roseus.

Les cellules de l'invention peuvent aussi être immobilisées, ce qui permet d'obtenir une production constante et prolongée de l'hémoglobine. La séparation de l'hémoglobine et la biomasse végétale est aussi facilitée. Comme méthode d'immobilisation, on peut citer l'immobilisation en billes d'alginate, d'agar, à l'intérieur de mousse de polyuréthane, ou bien encore dans des fibres creuses.

Les cellules de l'invention peuvent également être des cultures de racines. Les racines cultivées in vitro, en milieu liquide, sont nommées "Hairy roots", ce sont des racines transformées par la bactérie Agrobacterium rhizogenes.

Au lieu de produire l'hémoglobine de l'invention par culture de cellules végétales, on peut régénérer des plantes chimériques ou transgéniques à partir d'explants transformés, en ayant recours à des techniques connues en soi.

Comme plantes appropriées, on peut citer les Angiospermes comprenant les monocotylédones et les dicotylédons. Plus particulièrement, on peut citer le tabac, les espèces appartenant aux familles botaniques telles que les légumineuses (par exemple les haricots, pois, etc...), les crucifères (par exemple les choux, radis, colza, etc...), les solanacées (par exemple les tomates, pommes de terre, etc...), les cucurbitacées (par exemple le melon), les chénopodiacées (par exemple la betterave potagère), les ombellifères (par exemple les carottes, céleris, etc...). On peut également citer les céréales telles que le blé, le maïs, l'orge, le triticale et le riz, et les

oléagineux tels que le tournesol et le soja. Le tabac, la pomme de terre, la tomate et le maïs sont particulièrement préférés. Pour la pomme de terre, l'expression a lieu de préférence dans les tubercules.

L'invention concerne également les graines des plantes transgéniques capables de produire l'hémoglobine ainsi que leurs descendances.

L'invention vise également les protéines héminiques susceptibles d'être obtenues par le procédé de l'invention, en particulier les protéines héminiques capables de fixer de manière réversible l'oxygène, par exemple les hémoglobines et dérivés de celles-ci.

Les hémoglobines de l'invention sont capables de fixer l' $O_2$  de façon réversible avec une affinité ( $P_{50}$ ) de préférence proche des valeurs physiologiques (37°C), pH 7,40). L'affinité de la molécule pour l' $O_2$  est exprimée en  $P_{50}$  : c'est-à-dire la pression partielle d' $O_2$  lorsque l'hémoglobine ou ses dérivés est saturée à 50%. La  $P_{50}$  est mesurée selon les techniques habituelles, par exemple au moyen d'un analyseur qui mesure le pourcentage de saturation en  $O_2$  en fonction de la pression en  $O_2$  (Kister et al., 1987). Normalement, les hémoglobines de l'invention présentent une vitesse d'auto-oxydation acceptable pour minimiser la formation de méthémoglobine impropre au transport de l' $O_2$ . Cette caractéristique peut être mesurée par le spectre d'absorption.

De préférence, les hémoglobines de l'invention sont des tétramères, de préférence alpha<sub>2</sub> bêta<sub>2</sub>, bêta<sub>4</sub>, ou éventuellement des tétramères de sous-unités chimériques  $\alpha/\beta$  (Dumoulin et al., 1994, Art. Cells, Blood Subst., and Immob. Biotech., 22, 733-738) ou des multiples de 4 sous-unités. La taille physique du

complexe doit être au moins celle du tétramère afin d'éviter sa filtration par les reins.

Les protéines héminiques de l'invention peuvent être utilisées dans de nombreuses applications pharmaceutiques, cosmétiques ou industrielles. L'invention concerne en particulier des compositions pharmaceutiques comprenant une ou plusieurs protéine(s) héminique(s) selon l'une quelconque des revendications 15 à 23 en association avec un excipient acceptable du point de vue physiologique.

Dans le domaine pharmaceutique, toutes les conditions nécessitant une amélioration du transport de l'oxygène peuvent être traitées par les hémoglobines de l'invention, ces conditions comprennent les suivantes :

- hémorragies aigües ou chroniques,
- états de chocs,
- angioplasties coronariennes ou sylviennes,
- traitements des tumeurs solides, sensibilisation à la gamma-thérapie,
- conservation d'organes avant greffe et pendant le transport,
- hémopathies malignes.

Les hémoglobines de l'invention sont normalement utilisées sous forme d'injection dans des solutions éventuellement stabilisées en ce qui concerne la forme tétramérique du complexe (par exemple, addition de pyridoxal phosphate ou diaspirine) en ce qui concerne l'auto-oxydation. Il est également possible d'utiliser des suspensions d'hémoglobine greffée sur un support afin d'augmenter la durée de vie dans la circulation. Le support peut être tout support conventionnel dans ce domaine, par exemple les polysaccharides.

Différents aspects de l'invention sont illustrés dans les figures :

- Figure 1 : Protoporphyrine III (IX) à fer,
- Figure 2 : Séquence du cDNA de l' $\alpha$ -globine humaine (423 paires de bases), et protéine correspondante,
- Figure 3 : Séquence du cDNA de la  $\beta$ -globine humaine (438 paires de bases), et protéine correspondante,
- Figure 4 : Dispositif expérimental pour photolyse éclair. Un laser à impulsion sert pour la photodissociation des ligands de l'Hb : (HbCO  $\rightarrow$  Hb + CO). Un deuxième faisceau optique, orienté à 90°, détecte les changements d'absorption en fonction de temps après dissociation.
- Figure 5 : Cinétiques de recombinaison bimoléculaire du CO à l'hémoglobine dans l'extrait de plante. Les deux phases correspondent aux deux états allostériques de l'Hb : R (rapide) et T (lente). Conditions : 0.1 atm CO, pH 6-6, 25°C, environ 50% dissociation, 0-1, 1 et 10  $\mu$ M Hb.
- Figure 6 : Cinétiques de recombinaison du CO à l'hémoglobine en fonction du pourcentage de dissociation (par variation de l'énergie laser). Les cinétiques sont sensibles au nombre de ligands dissociés (1 à 4). À fort niveau de dissociation, l'Hb (désoxy ou mono-ligandée) bascule vers la forme lente "T". À faible énergie laser, les tétramères (principalement avec trois ligands) restent dans la forme rapide "R".
- Figure 7 : Analyse par western blot de l'extrait des graines du tabac transgénique T26-22 transformé par le plasmide pBIOC59. Les extraits de graines (75  $\mu$ g de protéines) d'un tabac non-transformé (1) et d'un tabac transgénique (T26-22) (2), des marqueurs de poids moléculaire (3) et de l'HbA (50 ng) (4) sont séparés en électrophorèse SDS-PAGE 17% en

conditions réductrices. Le western blot est réalisé dans les conditions décrites au chapitre X.a. Les marqueurs de poids moléculaire et les globines  $\alpha$  et  $\beta$  sont indiqués.

- Figure 8 : Analyse par western blot des fractions obtenues lors de la purification partielle. Les protéines des fractions éluées de la Sephadryl S-100 (37  $\mu$ g), de la S-Sepharose (30  $\mu$ g) obtenues pendant la purification à partir des mélanges de graines témoins [respectivement (2) et FE-Témoin (4)] et de graines accumulant la rHb [respectivement (3) et FE-rHb (5)], de l'HbA (50 ng)(1) et des marqueurs de poids moléculaire (6) ont été séparés en électrophorèse SDS-PAGE 17% en conditions réductrices. Le western blot est effectué dans les conditions décrites dans le chapitre X.a. Les marqueurs de poids moléculaires et les globines  $\alpha$  et  $\beta$  sont indiqués.

- Figure 9 : Cinétique de recombinaison du CO à la fraction FE-rHb. La cinétique après photolyse éclair est caractéristique de l'Hb tétramérique normale. La fraction FE-Témoin provenant des plantes de contrôle donne un signal d'amplitude de 1 mDO soit environ 50 fois plus faible que celle observée pour la fraction FE-rHb (48 mDO).

- Figure 10 : Cinétique de recombinaison du CO à la fraction FE-rHb à différents niveaux d'intensité de laser. Des résultats similaires sont observés pour l'HbA (figure 6).

- Figure 11 : Démonstration de la liaison réversible de l'oxygène à l'échantillon FE-rHb. Comme les échantillons d'oxyhémoglobine ne donnent que des signaux faibles, nous avons utilisé pour ces mesures la technique du mélange d'une atmosphère CO et d'O<sub>2</sub>. Après photodissociation du CO, la phase rapide correspond à la liaison de l'oxygène. L'oxygène est

ensuite remplacé par le CO qui peut être photodissocié à nouveau. La figure montre également les cinétiques de recombinaison du CO du même échantillon équilibré sous 1 atm ou 0.1 atm CO.

#### EXEMPLES

##### I. CONSTRUCTION DE PLASMIDES BINAIRES D'EXPRESSION DE BASE PERMETTANT LA PRODUCTION DE PROTEINES RECOMBINANTES DANS LES FEUILLES DE TABAC.

L'expression de gènes dans les feuilles de tabac nécessite les séquences régulatrices suivantes :

1) le promoteur constitutif double 35S (pd35S) du CaMV (virus de la mosaïque du chou-fleur). Il correspond à une duplication des séquences activant la transcription situées en amont de l'élément TATA du promoteur 35S naturel (Kay et al., 1987).

2) la séquence terminatrice de transcription, terminateur polyA 35S, qui correspond à la région en 3' non codante de la séquence du virus à ADN bicaténaire circulaire de la mosaïque du chou-fleur produisant le transcript 35S (Franck et al., 1980).

Les constructions des différents plasmides via l'utilisation de techniques d'ADN recombinant (Sambrook et al., 1989) dérivent de pBIOC4. Ce plasmide binaire dérive de pGA492 (An, 1986) qui contient entre les bordures droite et gauche, issues du plasmide pTiT37 d'Agrobacterium tumefaciens, sur son ADN de transfert, les séquences suivantes :

- le promoteur constitutif du gène nos codant pour la nopaline synthase (Depicker et al., 1982),
- la séquence codante du gène nptII codant pour la néomycine phosphotransférase II (Berg et Berg,

1983) délétée de la région des 8 premiers codons dont le codon initiateur méthionine ATG et fusionnée à la séquence des 14 premiers codons de la séquence codante du gène nos (Depicker et al., 1982), la séquence codante du gène nos dépourvue de la région des 14 premiers codons, le terminateur nos (Depicker et al., 1982), un polylinker (HindIII-XbaI-SacI-HpaI-KpnI-ClaI-BglII) précédant le gène cat codant pour la chloramphénicol acétyltransférase (Close et Rodriguez, 1982) et les séquences terminatrices du gène 6 du plasmide pTiA6 d'Agrobacterium tumefaciens (Liu et al., 1993).

Pour éliminer la quasi-totalité de la séquence codante du gène cat, le plasmide pGA492 a été doublement digéré par SacI (site de restriction du polylinker) et par ScaI (site de restriction présent dans la séquence du gène cat) puis soumis à l'action de l'enzyme T4 DNA polymérase (New England Biolabs) selon les recommandations du fabricant. La ligation du plasmide modifié (20 ng) a été réalisée dans un milieu réactionnel de 10 µl contenant 1 µl de tampon T4 DNA ligase x 10 (Amersham) et 2,5 U d'enzyme T4 DNA ligase (Amersham) à 14°C pendant 16 heures. Les bactéries, E. coli DH5 $\alpha$  rendues préalablement compétentes, ont été transformées (Hanahan, 1983).

L'ADN plasmidique des clones obtenus, sélectionnés sur 12 µg/ml tétracycline, a été extrait selon la méthode de la lyse alcaline (Birnboim et Doly, 1979) et analysé par digestion enzymatique par des enzymes de restriction. Puis, le site de restriction HindIII de l'ADN plasmidique du clone retenu a été modifié en un site de restriction EcoRI à l'aide d'un adaptateur HindIII - EcoRI phosphorylé (Stratagene Cloning Systems). Pour réaliser cette modification, 500 ng d'ADN plasmidique du clone retenu

ont été digérés par HindIII, déphosphorylés par l'enzyme phosphatase alcaline d'int stin de veau (Boehringer Mannheim) selon les recommandations du fabricant et coprécipités en présence de 1500 ng d'ADN adaptateur HindIII-EcoRI, 1/10 volume d'acétate de sodium 3M pH4,8 et 2,5 volumes d'éthanol absolu à -80°C pendant 30 min. Après centrifugation à 12000g pendant 30 min., l'ADN précipité a été lavé à l'éthanol 70%, séché, repris dans 8 µl d'eau, porté à 65°C pendant 10 min., puis ligué en présence de 1 µl de tampon T4 DNA ligase x 10 (Amersham) et 2,5 U d'enzyme T4 DNA ligase (Amersham) à 14°C pendant 16 heures. Après inactivation de la T4 DNA ligase à 65°C pendant 10 min., le mélange réactionnel de ligation a été digéré par EcoRI, purifié par électrophorèse sur gel d'agarose 0,8%, électroélue (Sambrook et al., 1989), précipité en présence de 1/10 de volume de 3M acétate de sodium pH4,8 et de 2,5 volumes d'éthanol absolu à -80°C pendant 30 min., centrifugé à 12000g pendant 30 min., lavé à l'éthanol 70%, puis séché. Les bactéries, E. coli DH5 $\alpha$  rendues préalablement compétentes, ont été transformées (Hanahan, 1983). L'ADN plasmidique des clones obtenus, sélectionnés sur 12 µg/ml tétracycline, a été extrait selon la méthode de la lyse alcaline (Birnboim et Doly, 1979) et analysé par digestion enzymatique par HindIII et EcoRI notamment. Le plasmide binaire résultant, qui ne possède plus que les 9 derniers codons de la séquences codante du gène cat et dont le site EcoRI est unique, a été appelé pBIOC4.

a. CONSTRUCTION DU PLASMIDE BINAIRE D'EXPRESSION  
pBIOC21.

La cassette d'expression, constituée du promoteur pd35S et du terminateur polyA 35S, a été isolée à partir du plasmide pJIT163D. Le plasmide pJIT163D dérive du plasmide pJIT163 qui dérive lui-même du plasmide pJIT60 (Guerineau et Mullineaux, 1993). Le plasmide pJIT163 possède un codon ATG entre les sites HindIII et SalI du polylinker. Pour supprimer cet ATG et obtenir le plasmide pJIT163D, l'ADN plasmidique pJIT163 a été digéré doublement par HindIII et SalI, purifié par électrophorèse sur gel d'agarose 0,8%, électroéluié (Sambrook et al., 1989), précipité en présence de 1/10 de volume de 3M acétate de sodium pH4,8 et de 2,5 volumes d'éthanol absolu à -80°C pendant 30 min., centrifugé à 12000g pendant 30 min., lavé à l'éthanol 70%, séché, soumis à l'action de l'enzyme Klenow (New England Biolabs) selon les recommandations du fabricant, déprotéinisé par extraction avec 1 volume de phénol:chloroforme:alcool isoamylque (25:24:1) puis 1 volume de chloroforme:alcool isoamylque (24:1), précipité en présence de 1/10 de volume de 3M acétate de sodium pH4,8 et de 2,5 volumes d'éthanol absolu à -80°C pendant 30 min., centrifugé à 12000g pendant 30 min., lavé à l'éthanol 70%, séché, et enfin, ligué en présence de 1 µl de tampon T4 DNA ligase x 10 (Amersham) et 2,5 U d'enzyme T4 DNA ligase (Amersham) à 14°C pendant 16 heures. Les bactéries, E. coli DH5 $\alpha$  rendues préalablement compétentes, ont été transformées (Hanahan, 1983). L'ADN plasmidique des clones obtenus, sélectionnés sur 50 µg/ml ampicilline, a été extrait selon la méthode de la lyse alcaline (Birnboim et Doly, 1979) et analysé par digestion

enzymatique par des enzymes de restriction. Pour isoler la cassette d'expression constituée du promoteur pd35S et du terminateur polyA 35S (fragment SacI-XhoI), l'ADN plasmidique du clone pJIT163D retenu a été digéré par SacI et XhoI. Le fragment SacI-XhoI, portant la cassette d'expression, a été purifié par électrophorèse sur gel d'agarose 0,8%, électroélué (Sambrook et al., 1989), précipité en présence de 1/10 de volume de 3M acétate de sodium pH4,8 et de 2,5 volumes d'éthanol absolu à -80°C pendant 30 min., centrifugé à 12000g pendant 30 min., lavé à l'éthanol 70%, séché, puis soumis à l'action de l'enzyme Mung Bean nucléase (New England Biolabs) selon les recommandations du fabricant. Cet insert purifié (200 ng) a été cloné dans l'ADN plasmidique de pBIOC4 (20 ng) digéré par EcoRI, traité par l'enzyme Mung Bean nucléase et déphosphorylé par l'enzyme phosphatase alcaline d'intestin de veau (Boehringer Mannheim) selon les recommandations du fabricant. La réaction de ligation a été effectuée dans 20 µl en présence de 2 µl de tampon T4 DNA ligase x 10 (Amersham), de 2 µl de 50% polyéthylène glycol 8000 et de 5 U d'enzyme T4 DNA ligase (Amersham) à 14°C pendant 16 heures. Les bactéries, E. coli DH5 $\alpha$  rendues préalablement compétentes, ont été transformées (Hanahan, 1983). L'ADN plasmidique des clones obtenus, sélectionnés sur 12 µg/ml tétracycline, a été extrait selon la méthode de la lyse alcaline (Birnboim et Doly, 1979) et analysé par digestion enzymatique par des enzymes de restriction. Le plasmide résultant a été appelé pBIOC21.

b. CONSTRUCTION DU PLASMIDE BINAIRE DE CO-EXPRESSION  
pBIOC43.

Le plasmide binaire de co-expression permettra une expression de deux gènes dans le même vecteur binaire.

Le plasmide binaire de co-expression dérive de pBIOC21. Il contient deux cassettes d'expression constituées chacune d'un promoteur pd35S et d'un terminateur polyA 35S mais différent par le polylinker séparant le promoteur du terminateur. L'une des cassettes d'expression est celle de pBIOC21 déjà décrite au paragraphe I. a. L'autre cassette d'expression a été obtenue en remplaçant le polylinker HindIII-BamHI-SmaI-EcoRI de pJIT163D (décrit au paragraphe I. a.) par un adaptateur HindIII-EcoRI portant les sites de restriction PacI, AscI, MluI et HpaI. Cet adaptateur a été obtenu par renaturation des 2 oligodésoxynucléotides WD11 (5' AGC TGA TTA ATT AAG GCG CGC CAC GCG TTA AC 3') et WD12 (5' AAT TGT TAA CGC GTG GCG CGC CTT AAT TAA TC 3') qui sont complémentaires pour leurs 28 nucléotides 3' terminaux. Cent µM de chacun de ces deux oligodésoxynucléotides ont été préalablement phosphorylés par action de 10 U d'enzyme T4 polynucléotide kinase (New England Biolabs) dans un volume réactionnel total de 10 µl contenant 1 µl de tampon T4 polynucléotide kinase x 10 (New England Biolabs) et 3 µl d'ATP (95 mM). Les deux mélanges réactionnels ont été incubés à 37°C pendant 1 heure, puis à 65°C pendant 20 min. Ils ont été ensuite rassemblés et leur volume a été complété à 500 µl. Après extraction avec 1 volume de phénol:chloroforme:alcool isoamylique (25:24:1) et 1 volume de chloroforme:alcool isoamylique (24:1), 50 µl

d 3M acétate de sodium pH6.0 ont été ajoutés. Le mélange réactionnel a été incubé à 80°C pendant 10 min., puis refroidi lentement à température ambiante. L'ADN a ensuite été précipité en présence de 2,5 volumes d'éthanol absolu à -80°C pendant 30 min., centrifugé à 14000g à 4°C pendant 1 heure, lavé à l'éthanol 70%, centrifugé à 14000g à 4°C pendant 10 min., séché, repris dans 10 µl de H<sub>2</sub>O. Le fragment d'ADN HindIII-EcoRI a ensuite été cloné aux sites HindIII-EcoRI de l'ADN plasmidique pJIT163D préalablement déphosphorylé par l'enzyme phosphatase alcaline d'intestin de veau (New England Biolabs) selon les recommandations du fabricant. La réaction de ligation a été effectuée dans un volume réactionnel de 20 µl en présence de 1 U de T4 DNA ligase (Gibco-BRL) pour une concentration totale d'ADN de 8,5 nM avec un rapport molaire vecteur/insert de 1 et de 4 µl de tampon T4 DNA ligase x 5 (Gibco-BRL) à 25°C pendant 16 heures. Les bactéries, E. coli DH5α rendues préalablement compétentes, ont été transformées (Hanahan, 1985). L'ADN plasmidique des clones obtenus, sélectionnés sur 100 µg/ml ampicilline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique. Le clone résultant a été appelé pBIOC42. Sa validité a été vérifiée par séquençage à l'aide du kit "Sequenase Version 2.0 DNA Sequencing" commercialisé par United States Biochemical (USB) selon la méthode des didésoxynucléotides (Sanger et al., 1977). Les conditions réactionnelles suivent les indications du fabricant excepté pour la dénaturation et l'hybridation. Le milieu réactionnel contenant l'ADN plasmidique (0,5 à 1 pmol s), l'amorce oligonucléotidique (2 pmoles), 10% de DMSO et le tampon réactionnel x 1 (USB), est incubé à 100°C

pendant 10 min., puis refroidi brutalement à - 80°C dans la carboglace.

A partir de pBIOC42, le fragment d'ADN codant pour la cassette d'expression constituée du promoteur pd35S et du terminateur polyA 35S a été isolé par double digestion par SacI et XhoI. Il a été purifié par électrophorèse sur gel d'agarose 0,75%, puis soumis à l'action du kit "Geneclean II" commercialisé par BIO101 selon les indications du fabricant. Puis, ce fragment d'ADN a été inséré aux sites SacI et XhoI du plasmide pBCSK+ commercialisé par Stratagene et préalablement déphosphorylé par l'enzyme phosphatase alcaline d'intestin de veau (New England Biolabs) selon les recommandations du fabricant. La ligation a été réalisée dans un volume réactionnel de 20 µl en présence de 1 U de T4 DNA ligase (Gibco-BRL) pour une concentration totale d'ADN de 8,5 nM avec un rapport molaire vecteur/insert de 1 et de 4 µl de tampon T4 DNA ligase x 5 (Gibco-BRL) à 25°C pendant 16 heures. Les bactéries, E. coli DH5 $\alpha$  rendues préalablement compétentes, ont été transformées (Hanahan, 1985). L'ADN plasmidique des clones obtenus, sélectionnés sur 30 µg/ml de chloramphénicol, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique par des enzymes de restriction. Le plasmide résultant a été appelé pBIOC75.

A partir de pBIOC75, le fragment d'ADN portant la cassette d'expression constituée du promoteur pd35S et du terminateur polyA 35S a été isolée par digestion par KpnI. Il a été purifié par électrophorèse sur gel d'agarose 0,75%, puis soumis à l'action du kit "Geneclean II" commercialisé par BIO101 selon les indications du fabricant. Puis, ce fragment d'ADN a été ligué à l'ADN plasmidique de pBIOC21 digéré par

KpnI et déphosphorylé par l'enzyme phosphatase alcaline d'intestin de veau (New England Biolabs) selon les recommandations du fabricant. La ligation a été réalisée dans un volume réactionnel de 20 µl en présence de 1 U de T4 DNA ligase (Gibco-BRL) pour une concentration totale d'ADN de 8,5 nM avec un rapport molaire vecteur/insert de 1 et de 4 µl de tampon T4 DNA ligase x 5 (Gibco-BRL) à 25°C pendant 16 heures. Les bactéries, E. coli DH5 $\alpha$  rendues préalablement compétentes, ont été transformées (Hanahan, 1985). L'ADN plasmidique des clones obtenus, sélectionnés sur 12 µg/ml de tétracycline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique par des enzymes de restriction. Le plasmide résultant a été appelé pBIOC43.

## II. CONSTRUCTION DES GENES CHIMERIQUES CODANT POUR LES CHAINES DE GLOBINE, $\alpha$ ET $\beta$ , PERMETTANT UNE EXPRESSION DE L'HEMOGLOBINE HUMAINE RECOMBINANTE DANS LE CYTOPLASME DES FEUILLES DE TABAC.

Le plasmide alpha

JW101 renferme le cDNA de la chaîne  $\alpha$  de globine cloné dans le plasmide pMB9 comme décrit par Wilson et al. (1978).

Le phage M13mp10 cIIIFX bêta renferme le cDNA de la chaîne  $\beta$  de globine cloné dans le phage M13 mp10 comme décrit par Nagai et al. (1985). Dans cette construction, l'ADN codant pour la  $\beta$  globine a été inséré en 3' de la séquence codante pour la protéine CII du phage lambda, suivie de celle codant pour le tétrapeptide FX, formant un gène de fusion dans lequel le codon initiateur ATG de la  $\beta$  globine a été déleté.

a. CONSTRUCTION DU PLASMIDE pBIOC44 CONTENANT LE cDNA CODANT POUR L' $\alpha$  GLOBINE POUR UN ADRESSAGE CYTOPLASMIQUE.

Pour obtenir un adressage cytoplasmique de la chaîne  $\alpha$  de globine, le codon méthionine initiateur de la chaîne  $\alpha$  globine a été maintenu.

Le cDNA codant pour la chaîne  $\alpha$  globine à adressage cytoplasmique a été obtenu en trois étapes. Les deux premières étapes ont permis de supprimer le site interne HindIII (substitution d'un T en un C) en position 276 de la séquence codante tandis que la troisième étape réunit les 2 fragments du cDNA codant pour la chaîne  $\alpha$  globine recombinante.

La première étape a consisté en l'amplification des 95 premiers codons de la chaîne  $\alpha$  globine mature sur le plasmide alpha1pJW101 à l'aide des 2 oligodésoxynucléotides, WD13 (5' tacaagcttaaca ATG GTG CTG TCT CCg GCC GAC 3') et AD27 (5' CGG GTC CAC CCG GAG CTT GTG 3'). L'amorce WD13 apporte le site de restriction HindIII, la séquence aaca favorisant l'initiation de la traduction (Joshi, 1987) et précédant le codon initiateur ATG suivi des 6 premiers codons de la chaîne  $\alpha$  globine mature dont le quatrième (CCT) est substitué en CCg (mutation silencieuse) pour créer le site de restriction EagI. L'amorce AD27 permet la suppression du site de restriction HindIII par substitution du nucléotide T en C (position 276 de la séquence codante). L'amplification PCR a été réalisée dans 100  $\mu$ l de milieu réactionnel contenant 10  $\mu$ l de tampon Taq DNA polymérase x 10 (100 mM Tris-HCl pH8,4, 500 mM KCl et 20 mM MgCl<sub>2</sub>), 16  $\mu$ l du mélange de dNTP (1,25 mM dATP, 1,25 mM dCTP, 1,25 mM dGTP et 1,25 mM dTTP), 10  $\mu$ l de chacune des amores décrivées ci-dessus à 10  $\mu$ M, 10  $\mu$ l d'ADN matrice

(alpha $\beta$ pJW101) à 1 ng/ $\mu$ l et 0,5  $\mu$ l de Taq DNA polyméras à 5 U/ $\mu$ l (Perkin Elmer). Trente cycles comportant chacun 30 sec. de dénaturation à 97°C, 1 min. d'hybridation à 55°C et 2 min. d'elongation à 72°C ont été effectués dans l'appareil "Crocodile II" d'Appligène. Les fragments d'ADN amplifiés ont ensuite été purifiés par électrophorèse sur gel d'agarose 1,8% et action du kit "Geneclean II" commercialisé par BIO101 selon les indications du fabricant. Les fragments d'ADN amplifiés purifiés sont repris dans 20  $\mu$ l.

La deuxième étape a consisté en l'amplification des 54 derniers codons de la chaîne  $\alpha$  globine mature sur le plasmide alpha $\beta$ pJW101 à l'aide des 2 oligodésoxynucléotides, AD26 (5' CAC AAG CTC CGG GTG GAC CCG 3') et WD14 (5' gcgaattc TCA ACG GTA TTT GGA GGT CAG CAC 3'). L'amorce WD14 apporte le site de restriction EcoRI situé juste après le codon stop. L'amorce AD26 permet la suppression du site de restriction HindIII par substitution du nucléotide T en C (position 276 de la séquence codante). L'amplification PCR a été réalisée comme décrite dans la première étape. Le traitement des fragments d'ADN amplifiés a été effectué comme décrit dans la première étape.

La troisième étape a été l'amplification par PCR du cDNA complet codant pour la chaîne  $\alpha$  globine (142 codons dont l'ATG initiateur). Les deux types de fragments d'ADN amplifiés dans les première et deuxième étapes ont servi d'ADN matrice et les deux amorces utilisées ont été WD13 et WD14. L'amplification PCR a été réalisée comme décrite dans la première étape excepté que la température d'hybridation du cycle est de 60°C. Les fragments d'ADN amplifiés ont ensuite été extraits par de

l'éther saturé en H<sub>2</sub>O après avoir ajusté le volume à 500 µl avec du tampon TE (10 mM Tris-HCl pH8,0, 1 mM EDTA). Après extraction avec 1 volume de phénol:chloroforme:alcool isoamylique (25:24:1) et 1 volume de chloroforme:alcool isoamylique (24:1), les fragments d'ADN ont été précipités en présence de 1/10 de volume de 3M acétate de sodium pH6,0 et de 2 volumes d'éthanol absolu à - 80°C pendant 30 min., centrifugé à 14000g à 4°C pendant 30 min., lavé à l'éthanol 70%, centrifugé à 14000g à 4°C pendant 10 min., séché, repris dans 50 µl de H<sub>2</sub>O. Puis, 25 µl de ces fragments d'ADN ont été digérés doublement par HindIII et EcoRI, purifiés par électrophorèse sur gel d'agarose 1,8% et par action du kit "Geneclean II" (BIO101) et clonés aux sites HindIII et EcoRI du plasmide pNEB193 commercialisé par New England Biolabs, et préalablement déphosphorylé par l'enzyme phosphatase alcaline d'intestin de veau (New England Biolabs) selon les recommandations du fabricant. La ligation et la transformation ont été réalisées comme décrites au chapitre I. b. L'ADN plasmidique des clones obtenus, sélectionnés sur 100 µg/ml ampicilline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique. Le clone résultant a été appelé pBIOC44. La séquence nucléotidique du cDNA codant pour la chaîne α globine recombinante a été vérifiée par séquençage à l'aide du kit "Sequenase Version 2.0 DNA Sequencing" commercialisé par United States Biochemical (USB) comme décrit au chapitre I. b. Le séquençage a révélé deux mutations silencieuses situées au quarante-huitième nucléotide (C modifié en T) et au cinquante-quatrième (T modifié en C) de la séquence codante pour la chaîne α globine.

b. CONSTRUCTION DU PLASMIDE pBIOC45 CONTENANT LE cDNA CODANT POUR LA B GLOBINE POUR UN ADRESSAGE CYTOPLASMIQUE.

Pour obtenir un adressage cytoplasmique de la chaîne B de globine, le codon méthionine a été fusionné au premier codon de la chaîne B globine mature en maintenant la phase ouverte de lecture puisque l'ATG avait été déleté dans la construction M13mp10 cIIIFX bêta.

Le cDNA codant pour la chaîne B globine à adressage cytoplasmique a été obtenu par amplification PCR des 146 codons constituant la chaîne B globine mature sur le phage M13mp10 cIIIFX bêta à l'aide des 2 oligodésoxynucléotides, WD15 (5' gtcattaaatcca ATG GTG CAC CTG ACT CCT GAG GAG AAG TCg GCC GTT AC 3') et WD16 (5' aatgagctcgtaacgcgt TTA GTG ATA CTT GTG GGC CAG GGC 3'). L'amorce WD15 apporte le site de restriction PacI, la séquence aaca favorisant l'initiation de la traduction (Joshi, 1987) et le codon initiateur ATG suivi des 12 premiers codons de la chaîne B globine mature dont le neuvième (TCT) est substitué en TCg (mutation silencieuse) pour créer le site de restriction EagI. L'amorce WD16 apporte les sites de restriction MluI, HpaI et SacI placés après le codon stop. L'amplification PCR et le traitement des fragments d'ADN amplifiés ont été réalisées comme décrits dans la troisième étape du chapitre II. a. Puis, 25 µl de ces fragments d'ADN ont été digérés doublement par PacI et SacI, purifiés par électrophorèse sur gel d'agarose 1,8% et par action du kit "Geneclean II" (BIO101) et clonés aux sites PacI et SacI du plasmide pNEB193 commercialisé par New England Biolabs, préalablement déphosphorylé par l'enzyme phosphatase alcaline d'intestin de veau (New

England Biolabs) selon les recommandations du fabricant. La ligation et la transformation ont été réalisées comme décrites au chapitre I. b. L'ADN plasmidique des clones obtenus, sélectionnés sur 100 µg/ml ampicilline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique. Le clone résultant a été appelé pBIOC45. La séquence nucléotidique du cDNA codant pour la chaîne β globine recombinante a été vérifiée par séquençage comme décrit au chapitre I. b.

c. CONSTRUCTION DES PLASMIDES BINAIRES d'EXPRESSION,  
pBIOC 46 ET pBIOC47, ET DU PLASMIDE BINAIRE DE CO-  
EXPRESSION pBIOC49 POUR L'ADRESSAGE CYTOPLASMIQUE.

c.1. CONSTRUCTION DU PLASMIDE BINAIRE pBIOC46  
CONTENANT LE cDNA CODANT POUR L'α GLOBINE POUR UN  
ADRESSAGE CYTOPLASMIQUE.

A partir de pBIOC44, le fragment HindIII-EcoRI portant le cDNA codant pour la chaîne α globine à adressage cytoplasmique a été isolé par double digestion enzymatique par HindIII et EcoRI, purifié par électrophorèse sur gel d'agarose 1,8% et par action du kit "Geneclean II" (BIO101). Puis, ce fragment d'ADN a été ligué à l'ADN plasmidique de pBIOC21 doublement digéré par HindIII et EcoRI, et déphosphorylé par l'enzyme phosphatase alcaline d'intestin de veau (New England Biolabs) selon les recommandations du fabricant. La ligation et la transformation ont été réalisées comme décrites au chapitre I. b. L'ADN plasmidique des clones obtenus, sélectionnés sur 10 µg/ml tétracycline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique. Le clone résultant a été appelé pBIOC46. La séquence

nucléotidique du cDNA codant pour la chaîne  $\alpha$  globine recombinante a été vérifiée par séquençage comme décrit au chapitre I. b. L'ADN plasmidique du vecteur binaire pBIOC46 a été introduit par transformation directe dans la souche LBA4404 d'Agrobacterium tumefaciens selon le procédé de Holsters et al. (1978). La validité du clone retenu a été vérifiée par digestion enzymatique de l'ADN plasmidique introduit.

c.2. CONSTRUCTION DU PLASMIDE BINAIRE pBIOC47 CONTENANT LE cDNA CODANT POUR LA  $\beta$  GLOBINE POUR UN ADRESSAGE CYTOPLASMIQUE.

A partir de pBIOC45, le fragment HindIII-EcoRI portant le cDNA codant pour la chaîne  $\beta$  globine à adressage cytoplasmique a été isolé par double digestion enzymatique par HindIII (digestion totale) et EcoRI (digestion partielle), purifié par électrophorèse sur gel d'agarose 1,8% et par action du kit "Geneclean II" (BIO101). Puis, ce fragment d'ADN a été ligué à l'ADN plasmidique de pBIOC21 doublement digéré par HindIII et EcoRI, et déphosphorylé par l'enzyme phosphatase alcaline d'intestin de veau (New England Biolabs) selon les recommandations du fabricant. La ligation et la transformation ont été réalisées comme décrites au chapitre I. b.

L'ADN plasmidique des clones obtenus, sélectionnés sur 10  $\mu$ g/ml tétracycline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique. Le clone résultant a été appelé pBIOC47. La séquence nucléotidique du cDNA codant pour la chaîne  $\beta$  globine recombinante a été vérifiée par séquençage comme décrit au chapitre I. b. L'ADN plasmidique du vecteur binaire pBIOC47 a été introduit par transformation

directe dans la souche LBA4404 d'Agrobacterium tumefaciens selon le procédé de Holsters et al. (1978). La validité du clone retenu a été vérifiée par digestion enzymatique de l'ADN plasmidique introduit.

c.3. CONSTRUCTION DU PLASMIDE BINAIRE DE CO-EXPRESSION pBIOC49 CONTENANT LES cDNA CODANT POUR LES GLOBINES,  $\alpha$  ET  $\beta$ , POUR UN ADRESSAGE CYTOPLASMIQUE.

Le fragment HindIII-EcoRI portant le cDNA codant pour la chaîne  $\alpha$  globine à adressage cytoplasmique a été isolé à partir de pBIOC44 décrit au chapitre II. c. 1., et ligué à l'ADN plasmidique de pBIOC43 doublement digéré par HindIII et EcoRI, et préalablement déphosphorylé par l'enzyme phosphatase alcaline d'intestin de veau (New England Biolabs) selon les recommandations du fabricant. La ligation et la transformation ont été réalisées comme décrite au chapitre I. b. L'ADN plasmidique des clones obtenus, sélectionnés sur 12  $\mu$ g/ml tétracycline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique. Le clone résultant a été appelé pBIOC48.

Le fragment PacI-MluI portant le cDNA codant pour la chaîne  $\beta$  globine à adressage cytoplasmique a été isolé à partir de pBIOC45 décrit au chapitre II. c. 2., et ligué à l'ADN plasmidique de pBIOC48 doublement digéré par PacI et MluI, et déphosphorylé par l'enzyme phosphatase alcaline d'intestin de veau (New England Biolabs) selon les recommandations du fabricant. La ligation et la transformation ont été réalisées comme décrites au chapitre I. b, excepté que la souche E. coli Sure tet<sup>r</sup> a été utilisée à la place de DH5 $\alpha$ . La souche Sure tet<sup>r</sup> dérive de la souche Sure (Stratagene) rendue sensible à la tétracycline par la perte de

l'épisom F'. L'ADN plasmidique des clones obtenus, sélectionnés sur 10 µg/ml tétracycline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique. Le clone résultant a été appelé pBIOC49.

La séquence nucléotidique des cDNA codant pour les chaînes α et β de globine recombinante a été vérifiée par séquençage comme décrit au chapitre I. b. L'ADN plasmidique du vecteur binaire pBIOC49 a été introduit par transformation directe dans la souche LBA4404 d'Agrobacterium tumefaciens selon le procédé de Holsters et al. (1978). La validité du clone retenu a été vérifiée par digestion enzymatique de l'ADN plasmidique introduit.

### III. CONSTRUCTION DES GENES CHIMERIQUES CODANT POUR LES CHAINES DE GLOBINE, α ET β, PERMETTANT UNE EXPRESSION DE L'HEMOGLOBINE HUMAINE RECOMBINANTE DANS LES MITOCHONDRIES DES FEUILLES DE TABAC.

Pour obtenir un adressage mitochondrial, la séquence codant pour le peptide transit du précurseur de la sous-unité β de l'ATPase-F1 mitochondriale de Nicotiana plumbaginifolia (ATG GCT TCT CGG AGG CTT CTC GCC TCT CTC CTC CGT CAA TCG GCT CAA CGT GGC GGC GGT CTA ATT TCC CGA TCG TTA GGA AAC TCC ATC CCT AAA TCC GCT TCA CGC GCC TCT TCA CGC GCA TCC CCT AAG GGA TTC CTC TTA AAC CGC GCC GTA CAG TAC) est fusionnée au premier codon de la séquence codant pour d'une part, la chaîne α globine mature (délétion de l'ATG initiateur) et d'autre part, la chaîne β globin mature (délétion de l'ATG initiateur) en maintenant les phases ouvertes de lecture.

La séquence codant pour la sous-unité B de l'ATPase F1 mitochondriale de Nicotiana plumbaginifolia est contenue dans le plasmide pTZ-catp2-1 fourni par Boutry. Ce plasmide correspond au plasmide pTZ18R renfermant l'ADNc (cNP10) comme décrit par Boutry et Chua (1985).

Le peptide transit N-terminal, composé de 54 acides aminés comme défini par Chaumont et al. (1994), a été utilisé lors de la réalisation des constructions.

a. CONSTRUCTION DU PLASMIDE pBIOC50 CONTENANT LE cDNA CODANT POUR L' $\alpha$  GLOBINE POUR UN ADRESSAGE MITOCHONDRIAL.

Pour obtenir un adressage mitochondrial de la chaîne  $\alpha$  de globine, la séquence codant pour le peptide transit du précurseur de la sous-unité B de l'ATPase-F1 mitochondriale de Nicotiana plumbaginifolia a été fusionnée au premier codon de la séquence codant pour la chaîne  $\alpha$  globine mature en maintenant la phase ouverte de lecture. La séquence de clivage entre les séquences du peptide transit et de la chaîne  $\alpha$  globine mature est Tyr - Val.

La séquence codant pour le peptide transit du précurseur de la sous-unité B de l'ATPase-F1 mitochondriale a été amplifiée par PCR sur le plasmide pTZ-catp2-1 à l'aide des 2 oligodésoxynucléotides, WD17 (5' cgcaagcttaaca ATG GCT TCT CGG AGG CTT CTC 3') et WD18 (5' tag aat tc GGC cGG AGA CAG CAC GTA CTG TAC GGC GCG GTT TAA G 3'). L'amorce WD17 apporte le site de restriction HindIII, la séquence aaca favorisant l'initiation de la traduction (Joshi, 1987) et les 7 premiers codons du peptide transit (dont l'ATG initiateur). L'amorce WD18 apporte le site de

restriction EcoRI, les 5 premiers codons de la séquence codant la chaîne  $\alpha$  globine mature (un site d'restriction EagI est créé par mutation silencieuse dans le quatrième codon (CCT modifié en CCg) et les 7 derniers codons de la séquence du peptide transit. L'amplification PCR et le traitement des fragments d'ADN amplifiés ont été réalisées comme décrits dans la troisième étape du chapitre II. a. Puis, ces fragments d'ADN ont été digérés doublement par HindIII et EagI, purifiés par électrophorèse sur gel d'agarose 1,8% et par action du kit "Geneclean II" (BIO101) et clonés aux sites HindIII et EagI du plasmide pBIOC44 décrit au chapitre II. a., préalablement purifié par électrophorèse sur gel d'agarose 0,75% et par le kit "Geneclean II". Le plasmide pBIOC44 a été déphosphorylé par l'enzyme phosphatase alcaline d'intestin de veau (New England Biolabs) selon les recommandations du fabricant. La ligation et la transformation ont été réalisées comme décrites au chapitre I. b. L'ADN plasmidique des clones obtenus, sélectionnés sur 100  $\mu$ g/ml ampicilline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique. Le clone résultant a été appelé pBIOC50. La séquence nucléotidique de ce gène chimérique résultant de la fusion traductionnelle entre la séquence codant pour le peptide transit et le cDNA codant pour la chaîne  $\alpha$  globine mature a été vérifiée par séquençage comme décrit au chapitre I. b. Le séquençage a révélé deux mutations silencieuses situées au dixième nucléotide (C modifié en A) et au cent-quarante-et-unième (C modifié en G) de la séquence codante pour le peptide transit.

b. CONSTRUCTION DU PLASMIDE pBIOC51 CONTENANT LE cDNA CODANT POUR LA  $\beta$  GLOBINE POUR UN ADRESSAGE MITOCHONDRIAL.

Pour obtenir un adressage mitochondrial de la chaîne  $\beta$  de globine, la séquence codant pour le peptide transit du précurseur de la sous-unité  $\beta$  de l'ATPase-F1 mitochondriale de Nicotiana plumbaginifolia a été fusionnée au premier codon de la séquence codant pour la chaîne  $\beta$  globine mature en maintenant la phase ouverte de lecture. La séquence de clivage entre les séquences du peptide transit et de la chaîne  $\beta$  globine mature est Tyr - Val.

La séquence codant pour le peptide transit du précurseur de la sous-unité  $\beta$  de l'ATPase-F1 mitochondriale a été amplifiée par PCR sur le plasmide pTZ-catp2-1 à l'aide des 2 oligodésoxynucléotides, WD19 (5' gtcattaaataaca ATG GCT TCT CGG AGG CTT CTC GCC TCT C 3') et WD20 (5' aatgagct C GGC cGA CTT CTC CTC AGG AGT CAG GTG CAC GTA CTG TAC GGC GCG GTT TAA G 3'). L'amorce WD19 apporte le site de restriction PacI, la séquence aaca favorisant l'initiation de la traduction (Joshi, 1987) et précédant les 9 premiers codons du peptide transit (dont l'ATG initiateur). L'amorce WD20 apporte le site de restriction SacI, les 10 premiers codons de la séquence codant la chaîne  $\beta$  globine mature (un site de restriction EagI est créé par mutation silencieuse dans le neuvième codon (TCT modifié en TCg)) et les 7 derniers codons de la séquence du peptide transit. L'amplification PCR et le traitement des fragments d'ADN amplifiés ont été réalisées comme décrits dans la troisième étape du chapitre II. a. Puis, ces fragments d'ADN ont été digérés doublement par PacI et EagI, purifiés par électrophorèse sur gel d'agarose 1,8% et par action du

kit "Geneclean II" (BIO101) et clonés aux sites PacI et EagI du plasmide pBIOC45 décrit au chapitre II. b., préalablement purifié par électrophorèse sur gel d'agarose 0,75% et par le kit "Geneclean II". Le plasmide pBIOC45 a été déphosphorylé par l'enzyme phosphatase alcaline d'intestin de veau (New England Biolabs) selon les recommandations du fabricant. La ligation et la transformation ont été réalisées comme décrites au chapitre I. b. L'ADN plasmidique des clones obtenus, sélectionnés sur 100 µg/ml ampicilline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique. Le clone résultant a été appelé pBIOC51. La séquence nucléotidique de ce gène chimérique résultant de la fusion traductionnelle entre la séquence codant pour le peptide transit et le cDNA codant pour la chaîne  $\beta$  globine mature a été vérifiée par séquençage comme décrit au chapitre I. b.

c. CONSTRUCTION DU PLASMIDE BINAIRE DE CO-EXPRESSION pBIOC53 CONTENANT LES cDNA CODANT POUR LES GLOBINES,  $\alpha$  ET  $\beta$ , POUR UN ADRESSAGE MITOCHONDRIAL.

Le fragment HindIII-EcoRI portant le cDNA codant pour la chaîne  $\alpha$  globine à adressage mitochondrial a été isolé à partir de pBIOC50 décrit au chapitre III. a., et ligué à l'ADN plasmidique de pBIOC43 doublement digéré par HindIII et EcoRI, et déphosphorylé par l'enzyme phosphatase alcaline d'intestin de veau (New England Biolabs) selon les recommandations du fabricant. La ligation et la transformation ont été réalisées comme décrite au chapitre I. b. L'ADN plasmidique des clones obtenus, sélectionnés sur 10 µg/ml tétracycline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par

digestion enzymatique. Le clone résultant a été appelé pBIOC52.

Le fragment PacI-MluI portant le cDNA codant pour la chaîne  $\beta$  globine à adressage mitochondrial a été isolé à partir de pBIOC51 décrit au chapitre III. b., et ligué à l'ADN plasmidique de pBIOC52 doublement digéré par PacI et MluI, et déphosphorylé par l'enzyme phosphatase alcaline d'intestin de veau (New England Biolabs). La ligation et la transformation ont été réalisées comme décrites au chapitre II.c.3 en utilisant la souche E. coli Sure tet<sup>r</sup>. L'ADN plasmidique des clones obtenus, sélectionnés sur 10  $\mu\text{g}/\text{ml}$  tétracycline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique. Le clone résultant a été appelé pBIOC53.

La séquence nucléotidique des cDNA codant pour les chaînes  $\alpha$  et  $\beta$  de globine recombinante permettant un adressage mitochondrial a été vérifiée par séquençage comme décrit au chapitre I. b. L'ADN plasmidique du vecteur binaire pBIOC53 a été introduit par transformation directe dans la souche LBA4404 d'Agrobacterium tumefaciens selon le procédé de Holsters et al. (1978). La validité du clone retenu a été vérifiée par digestion enzymatique de l'ADN plasmidique introduit.

**IV. CONSTRUCTION DES GENES CHIMERIQUES CODANT POUR LES CHAINES DE GLOBINE,  $\alpha$  ET  $\beta$ , PERMETTANT UNE EXPRESSION DE L'HEMOGLOBINE HUMAINE RECOMBINANTE DANS LES CHLOROPLASTES DES FEUILLES DE TABAC.**

Pour obtenir un adressage chloroplastique, la séquence codant pour le peptide transit du précurseur de la petite sous-unité de la ribulose 1,5-biphosphate carboxylase de Pisum sativum L. (ATG GCT TCT ATG ATA TCC TCT TCA GCT GTG ACT ACA GTC AGC CGT GCT TCT ACG GTG CAA TCG GCC GCG GTG GCT CCA TTC GGC GGC CTC AAA TCC ATG ACT GGA TTC CCA GTT AAG AAG GTC AAC ACT GAC ATT ACT TCC ATT ACA AGC AAT GGT GGA AGA GTA AAG TGC) est fusionnée au premier codon de la séquence codant pour d'une part, la chaîne  $\alpha$  globine mature (délétion de l'ATG initiateur) et d'autre part, la chaîne  $\beta$  globine mature (délétion de l'ATG initiateur) en maintenant les phases ouvertes de lecture.

Ce peptide transit N-terminal, composé de 57 acides aminés comme défini par Anderson et al. (1986), a été isolé à partir du plasmide pJIT117 (Guerineau et al., 1988) et utilisé lors de la réalisation des constructions.

**a. CONSTRUCTION DU PLASMIDE pBIOC55 CONTENANT LE cDNA CODANT POUR L' $\alpha$  GLOBINE POUR UN ADRESSAGE CHLOROPLASTIQUE.**

Pour obtenir un adressage chloroplastique de la chaîne  $\alpha$  de globine, la séquence codant pour le peptide transit du précurseur de la petite sous-unité de la ribulose 1,5-biphosphate carboxylase de Pisum sativum L. a été fusionnée au premier codon de la séquence codant pour la chaîne  $\alpha$  globine mature en maintenant la phase ouverte de lecture. La séquence de

clivage entre les séquences du peptide transit et de la chaîne  $\alpha$  globine mature est Cys - Val.

La séquence du peptide transit du précurseur de la petite sous-unité de la ribulose 1,5-biphosphate carboxylase a été amplifiée par PCR sur le plasmide pJIT117 à l'aide des 2 oligodésoxynucléotides, WD21 (5' cgcaagcttaaca ATG GCT TCT ATG ATA TCC TCT TCA GC 3') et WD22 (5' tag aat tC GGC cGG AGA CAG CAC GCA CTT TAC TCT TCC ACC ATT GC 3'). L'amorce WD21 apporte le site de restriction HindIII, la séquence aaca favorisant l'initiation de la traduction (Joshi, 1987) et les 8 premiers codons du peptide transit (dont l'ATG initiateur). L'amorce WD22 apporte le site de restriction EcoRI, les 5 premiers codons de la séquence codant la chaîne  $\alpha$  globine mature (un site de restriction EagI est créé par mutation silencieuse dans le quatrième codon (CCT modifié en CCg)) et les 7 derniers codons de la séquence du peptide transit. L'amplification PCR et le traitement des fragments d'ADN amplifiés ont été réalisés comme décrits dans la troisième étape du chapitre II. a.

Puis, ces fragments d'ADN ont été digérés doublement par HindIII et EcoRI et clonés aux sites HindIII et EcoRI du plasmide pNEB193 commercialisé par New England Biolabs. Le plasmide pNEB193 a été déphosphorylé comme décrit en II. a. La ligation et la transformation ont été réalisées comme décrites au chapitre I. b. L'ADN plasmidique des clones obtenus, sélectionnés sur 100  $\mu$ g/ml ampicilline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique. Le clone résultant a été appelé pBIOC54. La séquence nucléotidique de ce gène chimérique résultant de la fusion traductionnelle entre la séquence codant pour le peptide transit et le cDNA codant pour la chaîne  $\alpha$

globine mature a été vérifiée par séquençage comme décrit au chapitre I. b.

A partir du plasmide pBIOC54, le fragment HindIII-EagI, portant la séquence codant pour le peptide transit du précurseur de la petite sous-unité de la ribulose 1,5-biphosphate carboxylase et les 4 premiers codons de la chaîne  $\alpha$  globine mature, a été isolé par double digestion, HindIII (digestion totale) et EagI (digestion partielle). Ce fragment HindIII-EagI, purifié par électrophorèse sur gel d'agarose 1,8% et par action du kit "Geneclean II" (BIO101), a été cloné aux sites HindIII et EagI du plasmide pBIOC44 déphosphorylé comme décrit au chapitre II. a. La ligation et la transformation ont été réalisées comme décrites au chapitre I. b. L'ADN plasmidique des clones obtenus, sélectionnés sur 100  $\mu$ g/ml ampicilline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique. Le clone résultant a été appelé pBIOC55. La séquence nucléotidique de ce gène chimérique résultant de la fusion traductionnelle entre la séquence codant pour le peptide transit et le cDNA codant pour la chaîne  $\alpha$  globine mature a été vérifiée par séquençage comme décrit au chapitre I. b.

b. CONSTRUCTION DU PLASMIDE pBIOC57 CONTENANT LE cDNA CODANT POUR LA  $\beta$  GLOBINE POUR UN ADRESSAGE CHLOROPLASTIQUE.

Pour obtenir un adressage chloroplastique de la chaîne  $\beta$  de globine, la séquence codant pour le peptide transit du précurseur de la petite sous-unité de la ribulose 1,5-biphosphate carboxylase de Pisum sativum L. a été fusionnée au premier codon de la séquence codant pour la chaîne  $\beta$  globine mature en

maintenant la phase ouverte de lecture. La séquence de clivage entre les séquences du peptide transit et de la chaîne  $\beta$  globine mature est Cys - Val.

La séquence codant pour le peptide transit du précurseur de la petite sous-unité de la ribulose 1,5-biphosphate carboxylase a été amplifiée par PCR sur le plasmide pJIT117 à l'aide des 2 oligodésoxynucléotides, WD23 (5' gtcattaaatcaa ATG GCT TCT ATG ATA TCC TCT TCA GCT GTG 3') et WD24 (5' aatgagct C GGC cGA CTT CTC CTC AGG AGT CAG GTG CAC GCA CTT TAC TCT TCC ACC 3'). L'amorce WD23 apporte le site de restriction PacI, la séquence aaca favorisant l'initiation de la traduction (Joshi, 1987) et précédant les 10 premiers codons du peptide transit (dont l'ATG initiateur). L'amorce WD24 apporte le site de restriction SacI, les 10 premiers codons de la séquence codant la chaîne  $\beta$  globine mature (un site de restriction EagI est créé par mutation silencieuse dans le neuvième codon (TCT modifié en TCg)) et les 6 derniers codons de la séquence du peptide transit. L'amplification PCR et le traitement des fragments d'ADN amplifiés ont été réalisés comme décrits dans la troisième étape du chapitre II. a. Puis, ces fragments d'ADN ont été digérés doublement par PacI et SacI, purifiés par électrophorèse sur gel d'agarose 1,8% et par action du kit "Geneclean II" (BIO101) et clonés aux sites PacI et SacI du plasmide pNEB193 commercialisé par New England Biolabs. Le plasmide pNEB193 a été déphosphorylé comme décrit en II. a. La ligation et la transformation ont été réalisées comme décrites au chapitre I. b. L'ADN plasmidique des clones obtenus, sélectionnés sur 100  $\mu$ g/ml ampicilline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique. Le clone résultant a été appelé

pBIOC56. La séquence nucléotidique de ce gène chimérique résultant de la fusion traductionnelle entre la séquence codant pour le peptide transit et le cDNA codant pour la chaîne  $\beta$  globine mature a été vérifiée par séquençage comme décrit au chapitre I. b.

A partir du plasmide pBIOC56, le fragment PacI-EagI, portant la séquence du peptide transit du précurseur de la petite sous-unité de la ribulose 1,5-biphosphate carboxylase et les 9 premiers codons de la séquence codant la chaîne  $\beta$  globine mature, a été isolé par double digestion, PacI (digestion totale) et EagI (digestion partielle). Ce fragment PacI-EagI, purifié par électrophorèse sur gel d'agarose 1,8% et par action du kit "Geneclean II" (BIO101), a été cloné aux sites PacI et EagI du plasmide pBIOC45 déphosphorylé comme décrit au chapitre II. a. La ligation et la transformation ont été réalisées comme décrites au chapitre I. b. L'ADN plasmidique des clones obtenus, sélectionnés sur 100  $\mu$ g/ml ampicilline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique. Le clone résultant a été appelé pBIOC57. La séquence nucléotidique de ce gène chimérique résultant de la fusion traductionnelle entre la séquence codant pour le peptide transit et le cDNA codant pour la chaîne  $\beta$  globine mature a été vérifiée par séquençage comme décrit au chapitre I. b.

c. CONSTRUCTION DU PLASMIDE BINAIRE DE CO-EXPRESSION  
pBIOC59 CONTENANT LES cDNA CODANT POUR LES GLOBINES,  $\alpha$   
ET  $\beta$ , POUR UN ADRESSAGE CHLOROPLASTIQUE.

Le fragment HindIII-EcoRI portant le cDNA codant pour la chaîne  $\alpha$  globine à adressage chloroplastique a été isolé à partir de pBIOC55 décrit au chapitre IV. a., et ligué à l'ADN plasmidique de pBIOC43 doublement digéré par HindIII et EcoRI, et déphosphorylé comme décrit en II. a. La ligation et la transformation ont été réalisées comme décrite au chapitre I. b. L'ADN plasmidique des clones obtenus, sélectionnés sur 10  $\mu$ g/ml tétracycline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique. Le clone résultant a été appelé pBIOC58.

Le fragment PacI-MluI portant le cDNA codant pour la chaîne  $\beta$  globine à adressage chloroplastique a été isolé à partir de pBIOC57 décrit au chapitre IV. b., et ligué à l'ADN plasmidique de pBIOC58 doublement digéré par PacI et MluI, et déphosphorylé comme décrit en II. a. La ligation et la transformation ont été réalisées comme décrites au chapitre II.c.3 en utilisant la souche E. coli Sure tet<sup>r</sup>. L'ADN plasmidique des clones obtenus, sélectionnés sur 10  $\mu$ g/ml tétracycline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique. Le clone résultant a été appelé pBIOC59.

La séquence nucléotidique des cDNA codant pour les chaînes  $\alpha$  et  $\beta$  de globine recombinante permettant un adressage chloroplastique a été vérifiée par séquençage comme décrit au chapitre I. b. L'ADN plasmidique du vecteur binaire pBIOC59 a été introduit par transformation directe dans la souche LBA4404

d'Agrobacterium tumefaciens selon le procédé de Holsters et al. (1978). La validité du clone retenu a été vérifiée par digestion enzymatique de l'ADN plasmidique introduit.

**V. CONSTRUCTION DES GENES CHIMERIQUES CODANT POUR LES CHAINES DE GLOBINE,  $\alpha$  ET  $\beta$ , PERMETTANT UNE EXPRESSION DE L'HEMOGLOBINE HUMAINE RECOMBINANTE POUR LA SECRETION DANS LES FEUILLES DE TABAC.**

Pour obtenir la sécrétion, la séquence codant pour le peptide signal (PS) de la sporamine A des racines tubérisées de la patate douce (Murakami et al., 1986 ; Matsuoka et Nakamura, 1991) (ATG AAA GCC TTC ACA CTC GCT CTC TTC TTA GCT CTT TCC CTC TAT CTC CTG CCC AAT CCA GCC CAT TCC), est fusionnée au premier codon de la séquence codant pour d'une part, la chaîne  $\alpha$  globine mature (délétion de l'ATG initiateur) et d'autre part, la chaîne  $\beta$  globine mature (délétion de l'ATG initiateur) en maintenant les phases ouvertes de lecture. Ce peptide signal de 23 acides aminés a été isolé à partir du plasmide pMAT103 (Matsuoka et Nakamura, 1991) et utilisé lors de la réalisation des constructions.

**a. CONSTRUCTION DU PLASMIDE pBIOC60 CONTENANT LE cDNA CODANT POUR L' $\alpha$  GLOBINE POUR LA SECRETION.**

Pour obtenir la sécrétion de la chaîne  $\alpha$  de globine, la séquence codant pour le peptide signal de la sporamine A de la patate douce a été fusionnée au premier codon de la chaîne  $\alpha$  de globine mature en maintenant la phase ouverte de lecture. La séquence de

clivage entre les séquences du peptide signal et de la chaîne  $\alpha$  de globine mature est Ser-Val.

La séquence codant pour le peptide signal (PS) de la sporamine A des racines tubérisées de patate douce a été amplifiée par PCR sur le plasmide pMAT103 à l'aide des 2 oligodésoxynucléotides, WD25 (5' cgcaagcttaaca ATG AAA GCC TTC ACA CTC GC 3') et WD26 (5' tagaattC GGC CGG AGA CAG CAC GGA ATG GGC TGG ATT GGG CAG G 3'). L'amorce WD25 apporte le site de restriction HindIII, la séquence aaca favorisant l'initiation de la traduction (Joshi, 1987) et les 6 premiers codons du peptide signal (dont l'ATG initiateur). L'amorce WD26 apporte le site de restriction EcoRI, les 5 premiers codons de la séquence codant la chaîne  $\alpha$  globine mature (un site de restriction EagI est créé par mutation silencieuse dans le quatrième codon (CCT modifié en CCg)) et les 7 derniers codons de la séquence du peptide signal. L'amplification PCR et le traitement des fragments d'ADN amplifiés ont été réalisées comme décrits dans la troisième étape du chapitre II. a. Puis, ces fragments d'ADN ont été digérés doublement par HindIII et EagI, purifiés par électrophorèse sur gel d'agarose 1,8% et par action du kit "Geneclean II" (BIO101) et clonés aux sites HindIII et EagI du plasmide pBIOC44 déphosphorylé décrit au chapitre II. a. La ligation et la transformation ont été réalisées comme décrites au chapitre I. b. L'ADN plasmidique des clones obtenus, sélectionnés sur 100  $\mu$ g/ml ampicilline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique. Le clone résultant a été appelé pBIOC60. La séquence nucléotidique de ce gène chimérique résultant de la fusion traductionnelle entre la séquence codant pour le peptide signal et le cDNA codant pour la chaîne  $\alpha$

globine mature a été vérifiée par séquençage comme décrit au chapitre I. b.

b. CONSTRUCTION DU PLASMIDE pBIOC61 CONTENANT LE cDNA CODANT POUR LA  $\beta$  GLOBINE POUR LA SECRETION.

Pour obtenir la sécrétion de la chaîne  $\beta$  de globine, la séquence codant pour le peptide signal de la sporamine A de la patate douce a été fusionnée au premier codon de la chaîne  $\beta$  de globine mature en maintenant la phase ouverte de lecture. La séquence de clivage entre les séquences du peptide signal et de la chaîne  $\beta$  de globine mature est Ser-Val.

La séquence codant pour le peptide signal (PS) de la sporamine A des racines tubérisées de patate douce a été amplifiée par PCR sur le plasmide pMAT103 à l'aide des 2 oligodésoxynucléotides, WD27 (5' gtcattaaatcca ATG AAA GCC TTC ACA CTC GC 3') et WD28 (5' aatgagct C GGC cGA CTT CTC CTC AGG AGT CAG GTG CAC GGA ATG GGC TGG ATT GGG CAG G 3'). L'amorce WD27 apporte le site de restriction PacI, la séquence aaca favorisant l'initiation de la traduction (Joshi, 1987) et les 6 premiers codons du peptide signal (dont l'ATG initiateur). L'amorce WD28 apporte le site de restriction SacI, les 10 premiers codons de la séquence codant la chaîne  $\beta$  globine mature (un site EagI est créé par mutation silencieuse dans le neuvième codon (TCT modifié en TCg)) et les 7 derniers codons de la séquence du peptide signal. L'amplification PCR et le traitement des fragments d'ADN amplifiés ont été réalisés comme décrits dans la troisième étape du chapitre II. a. Puis, ces fragments d'ADN ont été digérés doublement par PacI et EagI, purifiés par électrophorèse sur gel d'agarose 1,8% et par action du kit "Geneclean II" (BIO101) et clonés

aux sites PacI et EagI du plasmide pBIOC45 déphosphorylé décrit au chapitre II. b. La ligation et la transformation ont été réalisées comme décrites au chapitre I. b. L'ADN plasmidique des clones obtenus, sélectionnés sur 100 µg/ml ampicilline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique. Le clon résultant a été appelé pBIOC61. La séquence nucléotidique de ce gène chimérique résultant de la fusion traductionnelle entre la séquence codant pour le peptide signal et le cDNA codant pour la chaîne B globine mature a été vérifiée par séquençage comme décrit au chapitre I. b.

c. CONSTRUCTION DU PLASMIDE BINAIRE DE CO-EXPRESSION pBIOC63 CONTENANT LES cDNA CODANT POUR LES GLOBINES,  $\alpha$  ET  $\beta$ , POUR LA SECRÉTION.

Le fragment HindIII-EcoRI portant le cDNA codant pour la chaîne  $\alpha$  globine pour la sécrétion a été isolé à partir de pBIOC60 décrit au chapitre V. a., et ligué à l'ADN plasmidique de pBIOC43 doublement digéré par HindIII et EcoRI, et déphosphorylé comme décrit en II. a. La ligation et la transformation ont été réalisées comme décrites au chapitre II.c.3 en utilisant la souche E. coli Sure tet.. L'ADN plasmidique des clones obtenus, sélectionnés sur 10 µg/ml tétracycline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique. Le clone résultant a été appelé pBIOC62.

Le fragment PacI-MluI portant le cDNA codant pour la chaîne  $\beta$  globine pour la sécrétion a été isolé à partir de pBIOC61 décrit au chapitre V. b., et ligué à l'ADN plasmidique de pBIOC62 doublement digér' par PacI et MluI, et déphosphorylé comme décrit en II. a.

La ligation et la transformation ont été réalisées comme décrites au chapitre I. b. L'ADN plasmidique des clones obtenus, sélectionnés sur 10 µg/ml tétracycline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique. Le clone résultant a été appelé pBIOC63.

La séquence nucléotidique des cDNA codant pour les chaînes  $\alpha$  et  $\beta$  de globine pour la sécrétion a été vérifiée par séquençage comme décrit au chapitre I. b. L'ADN plasmidique du vecteur binaire pBIOC63 a été introduit par transformation directe dans la souche LBA4404 d'Agrobacterium tumefaciens selon le procédé de Holsters et al. (1978). La validité du clone retenu a été vérifiée par digestion enzymatique de l'ADN plasmidique introduit.

#### VI. CONSTRUCTION DES GENES CHIMERIQUES CODANT POUR LES CHAINES DE GLOBINE, $\alpha$ ET $\beta$ , PERMETTANT UNE EXPRESSION DE L'HEMOGLOBINE HUMAINE RECOMBINANTE DANS LE RETICULUM ENDOPLASMIQUE DES FEUILLES DE TABAC.

La séquence codant pour le signal KDEL (Lys-Asp-Glu-Leu) placée à l'extrémité C-terminale des chaînes  $\alpha$  et  $\beta$  de globine en amont du codon stop combinée à la présence de la séquence codant pour le peptide signal N-terminal (PS) de la sporamine A des racines tubérisées de patate douce permet un adressage dans le réticulum endoplasmique.

a. CONSTRUCTION DU PLASMIDE pBIOC65 CONTENANT LE cDNA CODANT POUR L'α GLOBINE PERMETTANT LA RETENTION DANS LE RETICULUM ENDOPLASMIQUE.

Pour obtenir une rétention dans le réticulum endoplasmique, la séquence codant pour le signal KDEL (5' aaa gag cta 3') a été placée avant le premier codon stop (TGA) de la chaîne α globine mature en maintenant la phase ouverte de lecture.

Le plasmide contenant le cDNA codant pour la chaîne α globine qui renferme la séquence codant pour le signal KDEL placée avant son premier codon stop a été obtenu en suivant les mêmes étapes que pour la fabrication du plasmide pBIOC44 décrit en II. a., excepté que l'amorce WD29 (5' gcgaattc TCA tag ctc atc ttt ACG GTA TTT GGA GGT CAG CAC 3') remplace l'amorce WD14. L'amorce WD29 apporte le site de restriction EcoRI et la séquence KDEL situés respectivement après et avant le codon stop.

Le plasmide résultant obtenu a été appelé pBIOC64. La séquence nucléotidique du gène chimérique entre le cDNA codant pour la chaîne α globine et la séquence codant pour le signal KDEL a été vérifiée par séquençage comme décrit au chapitre I. b.

Puis, le plasmide pBIOC64 a été modifié comme décrit en V. a. par fusion traductionnelle avec le peptide signal de la sporamine A des racines tubérisées de patate douce pour conduire au plasmide pBIOC65 permettant un adressage dans le réticulum endoplasmique. La séquence nucléotidique du gène chimérique entre la séquence codant pour le peptide signal, le cDNA codant pour la chaîne α globine mature et la séquence codant pour le signal KDEL a été vérifiée par séquençage comme décrit au chapitre I. b.

La séquenc de clivage entre les séquences du peptid signal et de la chaîne  $\alpha$  globine mature est Ser - Val.

b. CONSTRUCTION DU PLASMIDE pBIOC67 CONTENANT LE cDNA CODANT POUR LA  $\beta$  GLOBINE PERMETTANT LA RETENTION DANS LE RETICULUM ENDOPLASMIQUE.

Pour obtenir une rétention dans le reticulum endoplasmique, la séquence codant pour le signal KDEL (5' aaa gat gag cta 3') a été placée avant le premier codon stop (TAA) de la chaîne  $\beta$  globine mature en maintenant la phase ouverte de lecture.

Le plasmide contenant le cDNA codant pour la chaîne  $\beta$  globine qui renferme la séquence codant pour le signal KDEL avant son premier codon stop a été obtenu en suivant les mêmes étapes que pour la fabrication du plasmide pBIOC45 décrit en II. b., excepté que l'amorce WD30 (5' aatgagctcgtaacgcgt TTA tag ctc atc ttt GTG ATA CTT GTG GGC CAG GGC 3') remplace l'amorce WD16. L'amorce WD30 apporte les sites de restriction MluI, HpaI et SacI, et la séquence KDEL placés respectivement après et avant le codon stop.

Le plasmide résultant obtenu a été appelé pBIOC66. La séquence nucléotidique du gène chimérique entre le cDNA codant pour la chaîne  $\beta$  globine et la séquence codant pour le signal KDEL a été vérifiée par séquençage comme décrit au chapitre I. b.

Puis, le plasmide pBIOC66 a été modifié comme décrit en V. b. par fusion traductionnelle avec le peptide signal de la sporamine A des racines tubérisées de patate douce pour conduire au plasmide pBIOC67 permettant un adressage dans le réticulum endoplasmique. La séquence nucléotidique du gène chimérique entre la séquence codant pour le peptide

signal, le cDNA codant pour la chaîne  $\beta$  globine mature et la séquence codant pour le signal KDEL a été vérifiée par séquençage comme décrit au chapitre I. b. La séquence de clivage entre les séquences du peptide signal et de la chaîne  $\beta$  globine mature est Ser - Val.

c. CONSTRUCTION DU PLASMIDE BINAIRE DE CO-EXPRESSION pBIOC69 CONTENANT LES cDNA CODANT POUR LES GLOBINES,  $\alpha$  ET  $\beta$ , PERMETTANT LA RETENTION DANS LE RETICULUM ENDOPLASMIQUE.

Le fragment HindIII-EcoRI portant le cDNA codant pour la chaîne  $\alpha$  globine permettant la rétention dans le réticulum endoplasmique a été isolé à partir de pBIOC65 décrit au chapitre VI. a., et ligué à l'ADN plasmidique de pBIOC43 doublement digéré par HindIII et EcoRI, et déphosphorylé comme décrit en II. a. La ligation et la transformation ont été réalisées comme décrite au chapitre I. b. L'ADN plasmidique des clones obtenus, sélectionnés sur 10  $\mu$ g/ml tétracycline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique. Le clone résultant a été appelé pBIOC68.

Le fragment PacI-MluI portant le cDNA codant pour la chaîne  $\beta$  globine permettant la rétention dans le réticulum endoplasmique a été isolé à partir de pBIOC67 décrit au chapitre VI. b., et ligué à l'ADN plasmidique de pBIOC68 doublement digéré par PacI et MluI, et déphosphorylé comme décrit en II.a. La ligation et la transformation ont été réalisées comme décrites au chapitre II.c.3 en utilisant la souche E. coli Sure tet<sup>r</sup>. L'ADN plasmidique des clones obtenus, sélectionnés sur 10  $\mu$ g/ml tétracycline, a été extrait selon la méthode de la lyse alcaline (Stephen et al.,

1990) et analysé par digestion enzymatique. Le clone résultant a été appelé pBIOC69.

La séquence nucléotidique des cDNA codant pour les chaînes  $\alpha$  et  $\beta$  de globine permettant leur rétention dans le réticulum endoplasmique a été vérifiée par séquençage comme décrit au chapitre I. b. L'ADN plasmidique du vecteur binaire pBIOC69 a été introduit par transformation directe dans la souche LBA4404 d'Agrobacterium tumefaciens selon le procédé de Holsters et al. (1978). La validité du clone retenu a été vérifiée par digestion enzymatique de l'ADN plasmidique introduit.

#### VII. CONSTRUCTION DES GENES CHIMERIQUES CODANT POUR LES CHAINES DE GLOBINE, $\alpha$ ET $\beta$ , PERMETTANT UNE EXPRESSION DE L'HEMOGLOBINE HUMAINE RECOMBINANTE DANS LES VACUOLES DES FEUILLES DE TABAC.

Pour permettre l'adressage vacuolaire, la séquence codant pour le prépropeptide (PPS) de la sporamine A des racines tubérisées de la patate douce (Murakami et al., 1986 ; Matsuoka et Nakamura, 1991), qui correspond au peptide signal suivi de la séquence N-terminale d'adressage vacuolaire (ATG AAA GCC TTC ACA CTC GCT CTC TTC TTA GCT CTT TCC CTC TAT CTC CTG CCC AAT CCA GCC CAT TCC AGG TTC AAT CCC ATC CGC CTC CCC ACC ACA CAC GAA CCC GCC), est fusionnée au premier codon de la séquence codant pour d'une part, la chaîne  $\alpha$  globine mature (délétion de l'ATG initiateur) et d'autre part, la chaîne  $\beta$  globine mature (délétion de l'ATG initiateur) en maintenant les phases ouvertes de lecture. Ce prépropeptide de 37 acides aminés a été isolé à partir du plasmide pMAT103 (Matsuoka et

Nakamura, 1991) et utilisé lors de la réalisation des constructions.

Pour obtenir un adressage vacuolaire de la chaîne  $\alpha$  de globine, la séquence codant pour le prépropeptide de la sporamine A de la patate douce a été fusionnée au premier codon de la chaîne  $\alpha$  de globine mature en maintenant la phase ouverte de lecture. La séquence de clivage entre les séquences du peptide signal et de la chaîne  $\alpha$  de globine mature est Ala-Val.

a. CONSTRUCTION DU PLASMIDE pBIOC70 CONTENANT LE cDNA CODANT POUR L' $\alpha$  GLOBINE PERMETTANT UN ADRESSAGE VACUOLAIRE.

La séquence codant pour le prépropeptide N-terminal (PPS) de la sporamine A des racines tubérisées de patate douce a été amplifiée par PCR sur le plasmide pMAT103 à l'aide des 2 oligodésoxynucléotides, WD25 (5' cgcaagcttaaca ATG AAA GCC TTC ACA CTC GC 3') décrit en V. a. et WD31 (5' tagaattC GGC cGG AGA CAG CAC GGC GGG TTC GTG TGT GGT TG 3'). L'amorce WD31 apporte le site de restriction EcoRI, les 5 premiers codons de la séquence codant la chaîne  $\alpha$  globine mature (un site EagI est créé par mutation silencieuse dans le quatrième codon (CCT modifié en CCg)) et les 6 derniers codons de la séquence du prépropeptide N-terminal. L'amplification PCR et le traitement des fragments d'ADN amplifiés ont été réalisées comme décrits dans la troisième étape du chapitre II. a. Puis, ces fragments d'ADN ont été digérés doublement par HindIII et EagI, purifiés par électrophorèse sur gel d'agarose 1,8% et par action du kit "Geneclean II" (BIO101) et clonés aux sites HindIII et EagI du plasmide pBIOC44 déphosphorylé

d'crit au chapitre II. a. La ligation et la transformation ont été réalisées comme décrites au chapitre I. b. L'ADN plasmidique des clones obtenus, sélectionnés sur 100 µg/ml ampicilline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique. Le clone résultant a été appelé pBIOC70. La séquence nucléotidique du gène chimérique entre la séquence codant pour le prépropeptide et le cDNA codant pour la chaîne  $\alpha$  globine mature a été vérifiée par séquençage comme décrit au chapitre I. b.

b. CONSTRUCTION DU PLASMIDE pBIOC71 CONTENANT LE cDNA CODANT POUR LA  $\beta$  GLOBINE PERMETTANT UN ADRESSAGE VACUOLAIRE.

Pour obtenir un adressage vacuolaire de la chaîne  $\beta$  de globine, la séquence codant pour le prépropeptide de la sporamine A de la patate douce a été fusionnée au premier codon de la chaîne  $\beta$  de globine mature en maintenant la phase ouverte de lecture. La séquence de clivage entre les séquences du peptide signal et de la chaîne  $\beta$  de globine mature est Ala-Val.

La séquence codant pour le prépropeptide N-terminal (PPS) de la sporamine A des racines tubérisées de patate douce a été amplifiée par PCR sur le plasmide pMAT103 à l'aide des 2 oligodésoxynucléotides, WD27 (5' gtcattaaattaaca ATG AAA GCC TTC ACA CTC GC 3') décrit en V. b. et WD32 (5' aatgagct C GGC cGA CTT CTC CTC AGG AGT CAG GTG CAC GGC GGG TTC GTG TGT GGT TG 3'). L'amorce WD32 apporte le site de restriction SacI, les 10 premiers codons de la séquence codant la chaîne  $\beta$  globine mature (un site de restriction EagI est créé par mutation silencieuse dans le neuvième codon (TCT modifié en TCg)) et les 6

derniers codons de la séquence du prépropeptide N-terminal. L'amplification PCR et le traitement des fragments d'ADN amplifiés ont été réalisés comme décrits dans la troisième étape du chapitre II. a. Puis, ces fragments d'ADN ont été digérés doublement par PacI et EagI, purifiés par électrophorèse sur gel d'agarose 1,8% et par action du kit "Geneclean II" (BIO101) et clonés aux sites PacI et EagI du plasmide pBIOC45 déphosphorylé décrit au chapitre II. b. La ligation et la transformation ont été réalisées comme décrites au chapitre I. b. L'ADN plasmidique des clones obtenus, sélectionnés sur 100 µg/ml ampicilline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique. Le clone résultant a été appelé pBIOC71. La séquence nucléotidique du gène chimérique entre la séquence codant pour le prépropeptide et le cDNA codant pour la chaîne  $\beta$  globine mature a été vérifiée par séquençage comme décrit au chapitre I. b.

c. CONSTRUCTION DU PLASMIDE BINAIRE DE CO-EXPRESSION pBIOC73 CONTENANT LES cDNA CODANT POUR LES GLOBINES,  $\alpha$  ET  $\beta$ , PERMETTANT UN ADRESSAGE VACUOLAIRE.

Le fragment HindIII-EcoRI portant le cDNA codant pour la chaîne  $\alpha$  globine à adressage vacuolaire a été isolé à partir de pBIOC70 décrit au chapitre VII. a., et ligué à l'ADN plasmidique de pBIOC43 doublement digéré par HindIII et EcoRI, et déphosphorylé comme décrit en II.a. La ligation et la transformation ont été réalisées comme décrite au chapitre I. b. L'ADN plasmidique des clones obtenus, sélectionnés sur 10 µg/ml tétracycline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par

digestion enzymatique. Le clone résultant a été appelé pBIOC72.

Le fragment PacI-MluI portant le cDNA codant pour la chaîne  $\beta$  globine à adressage vacuolaire a été isolé à partir de pBIOC71 décrit au chapitre VII. b., et ligué à l'ADN plasmidique de pBIOC72 doublement digéré par PacI et MluI, et déphosphorylé comme décrit en II.a. La ligation et la transformation ont été réalisées comme décrites au chapitre II.c.3 en utilisant la souche E. coli Sure tet<sup>r</sup>. L'ADN plasmidique des clones obtenus, sélectionnés sur 10  $\mu\text{g}/\text{ml}$  tétracycline, a été extrait selon la méthode de la lyse alcaline (Stephen et al., 1990) et analysé par digestion enzymatique. Le clone résultant a été appelé pBIOC73.

La séquence nucléotidique des cDNA codant pour les chaînes  $\alpha$  et  $\beta$  de globine à adressage vacuolaire a été vérifiée par séquençage comme décrit au chapitre I. b. L'ADN plasmidique du vecteur binaire pBIOC73 a été introduit par transformation directe dans la souche LBA4404 d'Agrobacterium tumefaciens selon le procédé de Holsters et al. (1978). La validité du clone retenu a été vérifiée par digestion enzymatique de l'ADN plasmidique introduit.

#### VIII : OBTENTION DE PLANTES DE TABAC TRANSGENIQUES

Les plantes de tabac utilisées pour les expériences de transformation (Nicotiana tabacum var. PBD6) sont cultivées *in vitro* sur le milieu de base de Murashige et Skoog (1962) additionné des vitamines de Gamborg et al. (1968, Sigma référence M0404), de saccharose à 20g/L et d'agar (Merck) à 8g/L. Le pH du milieu est ajusté à 5,8 avec une solution de potasse avant autoclavage à 120°C pendant 20 min. Les

plantules de tabac sont repiquées par bouture des entre-noeuds tous les 30 jours sur ce milieu de multiplication MS20.

Toutes les cultures in vitro sont réalisées en enceinte climatisée, dans les conditions définies ci-dessous :

- Intensité lumineuse de  $30\mu\text{E}.\text{m}^{-2}.\text{s}^{-1}$  ; photopériode de 16h ;
- Thermopériode de 26°C le jour, 24°C la nuit.

La technique de transformation utilisée est dérivée de celle de Horsch et al. (1985).

Une préculture d'*Agrobacterium tumefaciens* souche LBA4404 contenant les plasmides pBIOC46 ou pBIOC47 ou pBIOC49 ou pBIOC53 ou pBIOC59 est réalisée durant 48h à 28°C sous agitation, dans du milieu LB additionné des antibiotiques adéquats (rifampicine et tétracycline). La préculture est ensuite diluée au 50ème dans le même milieu et cultivée dans les mêmes conditions. Après une nuit, la culture est centrifugée (10 min, 3000 g), les bactéries sont reprises dans un volume équivalent de milieu MS30 (30g/L saccharose) liquide et cette suspension est diluée au 10ème.

Des explants d'environ 1cm<sup>2</sup> sont découpés à partir des feuilles des plantules décrites ci-dessus. Ils sont ensuite mis au contact de la suspension bactérienne pendant 1h, puis séchés rapidement sur papier filtre et placés sur un milieu de coculture (MS30 solide).

Après 2 jours, les explants sont transférés en boîtes de Pétri sur le milieu de régénération MS30, contenant un agent sélectif, la kanamycine (200mg/L), un bactériostatique, l'augmentin (400mg/L) et les hormones nécessaires à l'induction de bourgeons (BAP, 1mg/L et ANA, 0,1mg/L). Un repiquage des explants est

effectué sur le même milieu après 2 semaines de culture. Après 2 nouvelles semaines, les bourgeons sont repiqués en boîtes de Pétri sur le milieu de développement composé du milieu MS20 additionné de kanamycine et d'augmentin. Après 15 jours, les bourgeons sont repiqués en pots sur le même milieu dont la concentration en kanamycine a été diminuée de moitié. L'enracinement prend environ 20 jours, au terme desquels les plantules peuvent être clonées par bouture d'entre-noeuds *in vitro* ou sorties en serre.

#### IX : EXTRACTION ET PURIFICATION PARTIELLES DE PROTEINES RECOMBINANTES A PARTIR DE FEUILLES DE TABAC

Cinquante grammes de feuilles de tabac (poids frais) transformées sont broyés dans l'azote liquide puis mis à agiter durant 15min à 4°C dans 300ml de tampon tris-HCl 50mM pH 8 additionné d'EDTA 1mM, de β-mercaptopropanoïde 1mM et de polyvinylpyrrolidone (PVP, 10g/300ml). Le broyat est filtré sur miracloth puis centrifugé pendant 20min à 4°C à 10000 g. Le surnageant est à nouveau filtré sur miracloth. Les protéines sont alors précipitées durant 12 h à 4°C par une solution de sulfate d'ammonium à saturation. Après centrifugation 20 min à 10000 g, le culot est repris dans du tampon tris-HCl 50mM pH 8 additionné de DTT 1mM et d'EDTA 1mM et dialysé 2 fois 12 heures contre ce même tampon. Après dialyse, le rétentat est centrifugé puis filtré sur miracloth. Un dosage de protéines est également effectué selon la technique de Bradford (1976).

Première étape de purification : Equilibration en tampon phosphate 10mM pH 6,7-EDTA 1 mM par passage sur une résine Sephadex G25 puis dépôt sur une résin échangeuse d'ions (CM cellulose) équilibrée en tampon

phosphate 10 mM pH 6,7 EDTA 1 mM. Lavage par 4 volumes de ce même tampon puis élution par un gradient linéaire de  $\text{Na}_2\text{HPO}_4$  10 mM pH 6,7 EDTA 1 mM à  $\text{Na}_2\text{HPO}_4$  100 mM pH 6,7 EDTA 1 mM.

Deuxième étape de purification : Equilibration en tampon Tris-HCl 10 mM pH 8,4-EDTA 1 mM par passage sur une résine Sephadex G25 puis dépôt sur une résine échangeuse d'ions DEAE-Sephacel équilibrée en tampon Tris-HCl 10 mM pH 8,4 EDTA 1 mM. Lavage par 4 volumes de ce même tampon puis élution par un tampon  $\text{KH}_2\text{PO}_4$  20 mM pH 7,4. Les conditions de pH et de force ionique peuvent être modifiées selon la nature du variant d'hémoglobine.

#### Détection de l'hémoglobine

L'hémoglobine (Hb) se détecte grâce à son chromophore, l'hème, qui lui donne sa couleur caractéristique. A faible concentration et en présence d'autre chromophore ou molécule qui diffuse la lumière, le signal dû à l'Hb peut être masqué. Ce problème peut être surmonté grâce à une technique dynamique permettant de détecter la présence de l'Hb dans un système complexe.

Cette méthode repose sur des spectres différentiels correspondant à une transition entre deux formes de l'Hb et sur les propriétés de photodissociation des ligands comme  $\text{O}_2$  et CO (Gibson, 1956 ; Mardenet et al., 1994). La probabilité d dissociation est plus forte pour le CO, donc ce ligand est utilisé de préférence. La préparation des échantillons s'effectue en anaérobiose.

L'équipement expérimental se compose de deux sources de lumières : la première est une source pulsée (laser) qui dissocie les ligands, et la deuxième est une lampe continue qui permet d'observer

la recombinaison des ligands, grâce à un changement de l'intensité de lumière transmise (figures 4, 5 et 6). La photodissociation est efficace dans tout le domaine spectral visible ; notre système consiste en un laser YAG dont les impulsions ont une durée de 10 ns à 532 nm. La détection est plus sensible dans la bande de Soret (416 nm) ; nous avons choisi 436 nm proche de l'absorption maximale de la forme désoxy. Les changements d'intensité transmise se produisent tout d'abord dans un temps de l'ordre de la nanosecond (phase géminée) puis se poursuivent en quelques millisecondes (phase bimoléculaire). Nous nous intéressons particulièrement à cette seconde phase qui reflète les transitions allostériques de l'Hb (figures 5 et 6). Les études cinétiques, rapides et réversibles, permettent d'obtenir de nombreuses données et donc une indication fiable de l'état de l'Hb quant à son fonctionnement normal, physiologique.

La préparation des échantillons est réalisée comme décrite ci-après. Les feuilles de tabac (20 g) sont broyées dans de l'azote liquide, puis le broyat est mélangé à 60 ml de la solution d'extraction (Tris-HCl 25 mM pH7,5,  $\beta$ -mercaptoéthanol 10 mM, EDTA 1 mM). L'homogénat est centrifugé à 10 000g à 4°C pendant 15 minutes. Le surnageant contenant les protéines solubles est prélevé. Le dosage des protéines est effectué selon la technique de Bradford (1976). A 1 ml d'extrait de protéines végétales (1 mg/ml), 32  $\mu$ l et 3.2  $\mu$ l d'une solution d'hémoglobine humaine concentrée (3.13 mg/ml) sont ajoutés pour obtenir respectivement des solutions contenant 100%, 10% et 1% d'hémoglobine par rapport aux protéines totales.

Les résultats obtenus sont les suivants :

Les cinétiques des échantillons équilibrés sous 0.1 atm CO pour trois concentrations en Hb : 100%, 10% et 1% des protéines totales présentes dans l'extrait à raison de 1 mg/ml ont été mesurées. Les courbes sont biphasiques, semblables à celles de l'Hb seule, et présentant une vitesse normale (de l'ordre de 1000/s) pour la phase rapide (figure 5). Les cinétiques sont similaires pour les deux concentrations, à l'exception de l'augmentation du bruit (signal) prévisible à faible concentration. Aucun signal n'a été observé pour l'extrait végétal en absence d'Hb, dans les mêmes conditions. Nous pouvons conclure que la cinétique de recombinaison du CO à l'HbA dans un extrait de feuilles de tabac est normale.

#### X : EXTRACTION ET PURIFICATION PARTIELLE DE L'HEMOGLOBINE RECOMBINANTE A PARTIR DE GRAINES DE TABAC.

Dans ce chapitre, on décrit les techniques utilisées, pour la détection par western blot, l'extraction et purification partielle et la mise en évidence de la fonctionnalité de l'hémoglobine recombinante produite dans les graines de tabacs transgéniques (rHb). Ceux-ci sont obtenus par transformation du plasmide de coexpression pBIOC 59 contenant les cDNA codant pour les globines,  $\alpha$  et  $\beta$ , permettant un adressage dans le chloroplaste.

a. DETECTION EN WESTERN BLOT DE L'HEMOGLOBINE RECOMBINANTE ACCUMULEE DANS LES GRAINES DE TABAC.

Soixante-quinze milligrammes de graines de tabac (poids frais) sont broyés dans l'azote liquide puis dans 600 µl de tampon Tris-HCl 25 mM pH 8 glacé additionné d'EDTA 1 mM, de DTT 1 mM et de PMSF 1 mM. Le broyat est transféré dans un tube Eppendorf et centrifugé à 4°C à 10000 g pendant 10 min. Le surnageant est alors concentré par ultrafiltration à l'aide des dispositifs micropure .45 et microcon 10 (Amicon). Le dosage des protéines est effectué selon la technique de Bradford (1976), en utilisant comme étalon l'albumine de sérum de boeuf (fraction V).

Les protéines sont séparées selon leur masse moléculaire apparente par électrophorèse en gel de polyacrylamide en présence de SDS d'après la méthode de Laemmli (Laemmli, 1970) en conditions réductrices. L'appareillage utilisé est le Mini-protean II (Bio-Rad). Le gel est constitué d'un gel de concentration (acrylamide 5 %, bis-acrylamide 0,17 %, Tris-HCl 63 mM pH 6,8, SDS 0,1 %) et d'un gel de séparation (acrylamide 17 %, bis-acrylamide 0,56 %, Tris-HCl 375 mM pH 8,8, SDS 0,1 %). Les échantillons protéiques sont préalablement dilués par 0,25 volume de solution de dépôt (Tris-HCl 200 mM pH 6,8, DTT 400 mM, glycérol 40 %, SDS 8 %, bleu de bromophénol 0,2 %), puis traités à 100°C pendant 5 min et enfin déposés sur le gel. L'électrophorèse est effectuée en tampon Tris-glycine-SDS (Tris 25 mM, glycine 250 mM, SDS 1 %) à 25 mA.

Après électrophorèse, les protéines sont transférées sur une membrane de nitrocellulose (BA 85, Schleicher & Schuell) par électrotransfert d'après la technique de Towbin et al. (1979). Le transfert est

réalisé à l'aide de l'appareil "mini trans blot module" (Bio-Rad) à 150 V pendant 90 min en présence de la solution de transfert (Tris 25 mM, glycine 192 mM, méthanol 20 %). La membrane est rincée pendant 5 min à température ambiante dans du tampon PBS 1X ( $\text{Na}_2\text{HPO}_4$  10,4 mM,  $\text{KH}_2\text{PO}_4$  3,2 mM, NaCl 116 mM), puis séchée.

La présence des chaînes de globine sur les western blot est détectée en utilisant comme anticorps primaire, un immunosérum anti-hémoglobine humaine de lapin (ref : H-4890, Sigma) et comme anticorps secondaire, un anticorps monoclonal anti-IgG de lapin couplé à la phosphatase alcaline (A-8025, Sigma). La révélation est effectuée en utilisant le substrat chromogène [5-bromo-4-chloro-3-indoyl phosphate/nitro blue tetrazolium (BCIP/NBT)].

La membrane est incubée sous agitation pendant 5 min dans une solution tampon TBST (Tris-HCl 10 mM pH 7,5, NaCl 150 mM, Tween 20 0,05 %), puis pendant au moins 30 min dans la même solution additionnée de 5% de lait écrémé en poudre (Régilait). Cette dernière solution est renouvelée, 1/5000 de volume de l'immunsérum anti-hémoglobine est ajouté et la membrane est incubée pendant au moins 2 heures. Elle est rincée 3 fois 5 min par de la solution TBST. L'incubation avec l'anticorps secondaire est effectuée pendant 1 heure avec l'anticorps monoclonal anti-IgG de lapin dilué à 1/10000 dans de la solution TBST. Ensuite, la membrane est de nouveau rincée 3 fois. L'activité phosphatase alcaline est révélée en incubant la membrane dans la solution de révélation (Tris-HCl 100 mM pH 9,5, NaCl 100 mM,  $\text{MgCl}_2$  5 mM, BCIP 330 µg/ml, NBT 165 µg/ml). La réaction est arrêtée par rinçage avec de l'eau.

La figure 7 représente l'analyse par western blot de la composition protéique des extraits de graines de tabacs transformés ou non par le plasmide pBIOC59. L'anticorps polyclonal reconnaît les deux chaînes de globine de l'hémoglobine normale adulte (HbA) séparées lors de l'électrophorèse SDS-PAGE. On observe que l'extrait protéique des graines de la plante transgénique T26-22 diffère de celui de la plante témoin par la présence de deux polypeptides dont la masse moléculaire apparente est similaire à celle des chaînes de globine de l'HbA et qui sont reconnus par l'anticorps. De plus, ils semblent être représentés de façon équimolaire. On peut donc dire que dans les graines, les transgènes codant pour les protéines de fusion peptide transit-globines  $\alpha$  et peptide transit-globine  $\beta$  sont exprimés ; le clivage du peptide transit serait correctement effectué, de sorte que s'accumulent les globines  $\alpha$  et  $\beta$ . Dans les graines de 11 plantes, sur les 20 tabacs transformés indépendamment par le plasmide pBIOC59, la présence des globines est détectée. Exprimé en équivalent d'HbA, le taux maximum d'environ 0,05% de rHb par rapport aux protéines totales solubles extraites est observé pour la plante T26-22. Cela a pu être estimé par analyse comparative en western de gammes de concentration d'HbA dans l'extrait protéique de graines de plante témoin.

b. EXTRACTION ET PURIFICATION PARTIELLE DE L'HEMOGLOBINE RECOMBINANTE A PARTIR DE GRAINES DE TABAC.

La purification partielle a été effectuée à partir d'un mélange des graines de tabacs transgéniques transformés par le plasmide pBIOC59 et exprimant la rHb.

Quinze grammes de graines de tabac (poids frais) sont broyés dans l'azote liquide puis dans 100 ml de tampon Tris-HCl 25mM pH 8 glacé additionné d'EDTA 1 mM, de DTT 1 mM et de PMSF 1 mM. Le broyat est filtré sur miracloth®, puis le filtrat est centrifugé à 4°C à 10000 g pendant 10 min. Le surnageant est d'abord saturé en monoxyde de carbone (CO), puis filtré avec un filtre 0,22 µm et enfin concentré par ultrafiltration à l'aide de dispositifs centriprep 10 (Amicon). Le concentrat est saturé en CO. Deux étapes chromatographiques successives sont réalisées (4°C) en suivant les absorbances à 280 nm (protéines) et 415 nm (hémoprotéines). (i) Le concentrat est préalablement filtré avec un filtre 0,22 µm, puis déposé sur colonne Sephadryl-S100 (Pharmacia) (2,1 cm x 90 cm) équilibré avec le tampon D ( $\text{Na}_2\text{HPO}_4$  9,12 mM,  $\text{NaH}_2\text{PO}_4$  20,88 mM, DTT 1 mM, EDTA 1 mM, pH 6,5). La fraction contenant la rHb est collectée, filtrée à travers un filtre 0,22 µm, puis saturée en CO et enfin concentrée comme précédemment. Soixante-cinq pour cent des protéines sont éliminées à cette étape. (ii) Ce concentrat est déposé sur la deuxième colonne, une S-sepharose fast-flow (Pharmacia) (1,1 cm x 10 cm) équilibrée avec du tampon D. Après lavage par 8 volumes de tampon D, un gradient de force ionique est appliqué (tampon D à tampon D contenant 500 mM NaCl). L'hémoglobine est éluee en un pic. Les fractions contenant ce pic sont

rassemblées et les protéines sont concentrées comme cela est décrit plus haut. Avant et après concentration les échantillons sont saturés en CO. Ce concentrat constitue la fraction enrichie en rHb appelée FE-rHb. Il ne reste plus dans cette fraction que 3 % des protéines de l'extrait. Pour disposer d'un témoin pour les analyses ultérieures, ce schéma de purification a été appliqué dans les mêmes conditions à un extrait obtenu à partir de 15 g de graines de tabacs n'exprimant pas la rHb, conduisant à l'obtention de la fraction appelée FE-Témoin.

La présence des globines  $\alpha$  et  $\beta$  dans ces fractions a été recherchée en utilisant la technique de western blot dans les conditions décrites au paragraphe X.a. La fraction FE-rHb contient bien de la rHb, ces deux polypeptides étant mis en évidence (figure 8).

#### C. MISE EN EVIDENCE DE LA FONCTIONNALITE DE L'HEMOGLOBINE RECOMBINANTE PAR PHOTOLYSE ECLAIR.

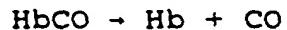
La mise en évidence de la fonctionnalité a été réalisée à partir de la fraction enrichie en rHb appelée FE-rHb, en utilisant comme témoin la fraction équivalente témoin FE-Témoin et de l'HbA.

Les expériences de contrôle où l'on a ajouté de l'HbA à l'extrait de plantes, ont montré des cinétiques de recombinaison biphasiques et des variations de la fraction lente selon l'énergie de l'éclair de lumière laser. Ces résultats démontrent que la fonction de l'HbA n'est pas altérée par les conditions de solvant utilisées.

Après photodissociation des ligands de l'Hb, la recombinaison bimoléculaire a lieu dans une échelle de temps de  $\mu$ s-ms (vitesse k-on). Bien que le ligand

physiologique naturel soit l'oxygène, les études décrites ont été faites avec du monoxyde de carbone (CO) qui donne un signal de photodissociation très supérieur à celui obtenu avec l'O<sub>2</sub> car le rendement est plus élevé. De même la différence des vitesses de recombinaison pour les deux conformations de l'Hb (R et T correspondant à des tétramères avec et sans ligand) est également plus élevée. Expérimentalement les échantillons sont équilibrés sous 0,1 atm CO qui donne les meilleures conditions d'observation des deux phases. Comme la réaction est réversible, on peut répéter la photodissociation ( $\gamma$ ) du même échantillon pour accumuler plusieurs courbes ce qui améliore beaucoup le rapport signal/bruit.

$\gamma$



L'observation d'une variation de l'amplitude de la recombinaison lente en fonction de la fraction de dissociation (par modification de l'énergie laser), démontre la présence d'une hémoglobine fonctionnelle.

Les plantes transgéniques reçoivent l'information génétique uniquement pour la synthèse de globine et non pour l'hème. En conséquence si l'Hb fonctionnelle (globines + hème) est exprimée dans les plantes, c'est qu'elle a capté l'hème in situ. D'autres hémoprotéines présentes dans les plantes peuvent donner un signal optique après photolyse éclair. Ces hémoprotéines ne donneront pas de signal si le fer héminique est sous forme ferrique qui ne lie pas les ligands CO et O<sub>2</sub>. Le CO et l'O<sub>2</sub> ne se lient de façon réversible que si l'atome de fer est sous forme ferreuse. Il est par

conséquent important de démontrer l'existence de processus cinétiques pour les deux phases et la variation des contributions relatives des deux phases due à des facteurs connus pour influencer la fonction de l'hémoglobine.

L'échantillon enrichi FE-rHb montre un signal de photodissociation du CO de 48 mDO (densité optique) et permet de réaliser certaines expériences à différents niveaux de dissociation ; ces expériences sont réalisées en l'absence de dithionite de sodium pour éviter toute contribution parasite due à la présence d'hémoprotéines. La même expérience, réalisée avec la fraction FE-Témoin a montré un signal de 1 mDO (figure 9).

Les résultats enregistrés à différents niveaux d'énergie de lumière laser sont montrés dans la Figure 10. Les courbes sont similaires à celle de l'HbA et montre l'existence d'une propriété caractéristique de l'hémoglobine à savoir la plus faible fraction de vitesse lente quand l'intensité de la lumière est diminuée de façon à obtenir une dissociation plus faible.

L'échantillon a été ensuite équilibré sous une atmosphère de CO. Comme attendu, les cinétiques de recombinaison sont désormais plus rapides. Pour l'hémoglobine en solution, la fraction lente est habituellement plus basse à fortes concentration de CO puisqu'il y a moins de temps disponible pour faire la transition R->T après dissociation. L'échantillon FE-rHb ne présente pas cet effet (figure 11).

Une autre méthode peut être utilisée pour étudier les vitesses d'association et de dissociation de l'oxygène. Le principe de cette méthode repose sur le fait suivant : bien que le CO ait une affinité environ 200 fois plus élevée que celle de l'oxygène, la

vitesse d'association du CO à l'Hb ligandée (R state) est environ 10 fois plus faible que pour l'oxygène. Un échantillon équilibré avec un mélange égal de CO et d'O<sub>2</sub> sera essentiellement sous forme de HbCO. On peut alors photodissocier le CO (avec un fort rendement) ce qui permet l'étude de la recombinaison de l'O<sub>2</sub>. Une phase terminale lente de l'ordre de 1s due au remplacement de l'oxygène par le CO fournit des informations sur la vitesse de dissociation (k-off). Seul l'échantillon FE-rHb révèle un signal de liaison de l'oxygène (figure 11).

Les études de la fraction FE-rHb par photolyse éclair ont montré :

- une recombinaison du CO biphasique avec des vitesses rapide et lente similaires à celles observées dans l'Hb A tétramérique ;
- une diminution de la fraction lente à énergie faible du laser comme pour l'Hb A ;
- une augmentation de la vitesse pour des concentrations de CO plus élevées comme pour l'Hb normale ;
- une liaison réversible de l'oxygène avec des vitesses on et off similaires à celles de l'Hb A normale ;

Nous concluons que l'hémoglobine recombinante produite dans les graines de tabac possède les propriétés de l'Hb A tétramérique dans tous les tests fonctionnels réalisés.

XI : CONSTRUCTION DE GENES CHIMERIQUES CODANT POUR LES CHAINES  $\alpha$  ET  $\beta$  DE L'HEMOGLOBINE HUMAINE ET PERMETTANT UNE EXPRESSION DANS LES SEMENCES DE MAIS.

CONSTRUCTION DES PLASMIDES CONTENANT UNE DES CHAINES  $\alpha$  OU  $\beta$  DE L'HEMOGLOBINE HUMAINE ET PERMETTANT UNE EXPRESSION CONSTITUTIVE OU UNE EXPRESSION DANS L'ALBUMEN DANS LES SEMENCES DE MAIS.

L'expression constitutive ou albumen spécifique dans les semences de maïs des séquences des chaînes  $\alpha$  et  $\beta$  de l'hémoglobine humaine a nécessité les séquences régulatrices suivantes :

un des trois promoteurs permettant une expression constitutive :

- promoteur actine de riz suivi de l'intron actine de riz (pAR-IAR) contenu dans le plasmide pAct1-F4 décrit par McElroy et al. (1991) ;

- promoteur constitutif double 35S (pd35S) du CaMV (virus de la mosaïque du chou-fleur). Il correspond à une duplication des séquences activant la transcription situées en amont de l'élément TATA du promoteur 35S naturel (Kay et al., 1987) ;

- le promoteur du gène de  $\gamma$ zéine de maïs (p $\gamma$ zéine) contenu dans le plasmide p $\gamma$ 63 (Reina et al., 1990). Le plasmide p $\gamma$ 63 résulte du clonage de p $\gamma$ zéine aux sites HindIII et XbaI d'un plasmide pUC18 renfermant, entre ses sites HindIII et EcoRI, la cassette d'expression "p35S-gus-tNOS" de pBI221 commercialisé par Clontech. Il permet une expression dans l'albumen des semences de maïs. Combiné à l'intron actine de riz, ce promoteur confère une expression est de type constitutive ;

un des deux terminateurs :

- la séquence terminatrice de transcription, terminateur polyA 35S, qui correspond à la région en 3' non codante de la séquence du virus à ADN bicaténaire circulaire de la mosaïque du chou-fleur produisant le transcript 35S (Franck et al., 1980) ;

- la séquence terminatrice de transcription, terminateur polyA NOS, qui correspond à la région en 3' non codante du gène de la nopaline synthase du plasmide Ti d'*Agrobacterium tumefaciens* souche à nopaline (Depicker et al., 1982).

Le type de vecteur utilisé dérive de pBSIISK+ (Stratagene). Chaque vecteur comporte une cassette d'expression, à savoir un des promoteurs, une des chaînes  $\alpha$  ou  $\beta$  de l'hémoglobine humaine et un des terminateurs. Des vecteurs comportant les deux cassettes d'expression de chacune des chaînes  $\alpha$  et  $\beta$  de l'hémoglobine humaine ont également été construits. Les clonages ont été réalisés selon les procédés usuels.

Références bibliographiques :

- Benesch & Kwong. *Hemoglobin* 1994, 18, 185-192.
- Birnboim & Doly. *Nucleic Acids Res.* 1979, 7, 1513-
- Boutry & Chua. *EMBO J.* 1985, 4, 2159-2165.
- Bradford. *Anal Biochem* 1976, 72, 248-254.
- Carter et al. *Mol Gen Genet* 1993, 241, 49-56.
- Chaumont et al. *Plant Mol Biol* 1994, 24, 631-641.
- Gamborg et al. *Exp Cell Res.* 1968, 50, 151-158.
- Guerineau et al. *Nucleic Acid Res.* 1988, 16 11380.
- Gibson. *J Physiol* 1956, 134, 123.
- Edelbaum. *J Interferon Res.* 1992, 12, 449-453.
- Hanahan. *J Mol Biol* 1983, 166, 557-
- Hanahan. In "DNA cloning volume I, a practical approach" (Ed: Glover D.M.) IRL Press, 1985, pp 109-135.
- Hiatt & Ma. *FEBS Lett* 1992, 307, 71-75.
- Hoffman et al. *Proc Natl Acad Sci USA* 1990, 87 : 8521-8525.
- Horsch et al. *Science* 1985, 227, 1229-1231.
- International Hemoglobin Information Center (1995) *Hemoglobin*, 19, 37-124.
- Jessen et al. *Meth Enzymol*, 1994, 231, 347-364.
- Joshi. *Nucleic Acid Res.* 1987, 15, 6643-6653.
- Kister et al. *J Biol Chem* 1987, 262, 12085-12091.
- Krebbers et al. *Plant Physiol*. 1988, 87, 859-866.
- Marden et al. *Meth Enzymol*. 1994, 232 71-86.
- Mason et al. *Proc Natl Acad Sci USA* 1992, 89, 11745-11749.
- Moloney. Int. Meeting of Production of Recombinant Proteins in Plants, Leicester 1994, page 36-38
- Murashige & Skoog. *Physiol Plantarum* 1962, 15, 473-497.
- Nagai & Thøgersen. *Meth Enzymol*. 1987, 153, 461-481.
- Nagai et al. *Proc Natl Acad Sci USA* 1985, 82, 7252-7255.
- Perutz. *Nature* 1970, 228, 726-739.
- Russel. Int. Meeting of Production of Recombinant Proteins in Plants, Leicester 1994, page 43
- Sambrook et al. *Molecular Cloning: A Laboratory Manual*. Second Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).
- Sanger et al. *Proc Natl Acad Sci USA* 1977, 74, 5463-5467
- Stephen et al. *Nucleic Acid Res.* 1990, 18, 7463-7464.
- Svab et al. *Proc Natl Acad Sci USA* 1990, 87, 8526-8530.
- Swanson et al. *Bio/Technology* 1991, 9, 57-61.
- Symons et al. *Bio/Technology* 1990, 8, 217-221.
- Vanderkerckhove et al. *Bio/Technology* 1989, 7, 929-932.
- Wagenbach et al. *Biotechnology* 1991, 9 : 57-61.
- Wilson et al. *Nucleic Acid Res.* 1978, 5, 563-581.

Kay R. et al., Science, 1987, 236 : 1299.

Franck A. et al., Cell, 1980, 21 : 285.

Depicker A. et al., J. Mol. Appl. Genet., 1982, 1 : 561.

Mc Elroy et al., Mol. Gen. Genet., 1991, 231 : 150.

Reina et al., N.A.R., 1990, 18 : 6426.

Dumoulin et al., Prot. Sci., 1996, 5 : 114-120.

Feng et al., J. Mol. Evol., 1985, 21 : 112-115.

Dumoulin et al., Art. Cells Blood Subs. Immob. Biotech., 1994, 22 : 733-738.

Looker et al., Nature, 1992, 356 : 258-260.

Laemmli, Nature, 1970, 227 : 680-685.

Towbin et al., Proc. Natl. Acad. Sci. USA, 1979, 76 : 4350-4354.

## REVENDICATIONS

1 - Procédé de production de protéines héminiques comprenant les étapes suivantes :

- i) introduction, dans des cellules végétales, d'une ou plusieurs molécule(s) d'acides nucléiques dont chacune comporte au moins une séquence codant pour un composant protéique d'une protéine héminique d'origine animale capable de fixer réversiblement l'oxygène, ou pour une variante ou une partie de ce composant protéique, et éventuellement une séquence codant pour un agent de sélection ;
- ii) sélection des cellules contenant l'acide nucléique codant pour le composant protéique de la protéine héminique ;
- iii) éventuellement, propagation des cellules transformées, soit en culture, soit par la régénération de plantes entières transgéniques ou chimériques ;
- iv) récupération, et éventuellement purification, d'une protéine héminique comprenant un complexe constitué de la protéine ou des protéines codées par le susdit acide nucléique et au moins un noyau porphyrique à fer, ou une pluralité de ces complexes.

2 - Procédé selon la revendication 1 caractérisé en ce que la protéine héminique est à chaîne polypeptidique unique, par exemple la myoglobine.

3 - Procédé selon la revendication 1 caractérisé en ce que la protéine héminique est un hétéro-oligomère, le ou les acide(s) nucléique(s) comportant les séquences codant pour chacune des différentes unités protéiques.

4 - Procédé selon la revendication 3, caractérisé en ce que la protéine héminique est l'hémoglobine humaine, ou un dérivé de celle-ci, le ou les acide(s)

nucléique(s) comprenant des séquences codant pour l' $\alpha$ - et  $\beta$ -globine, ou pour des variantes de l' $\alpha$ - ou de la  $\beta$ -globine, les variantes se différenciant de la séquence naturelle par une ou plusieurs substitution(s), délétion(s) ou insertion(s) d'acides aminés.

5 - Procédé selon la revendication 4 caractérisé en ce que les séquences codant pour les différentes unités protéiques,  $\alpha$ - et  $\beta$ -globine, sont comprises au sein de la même molécule d'acide nucléique.

6 - Procédé selon la revendication 4 caractérisé en ce que les séquences codant pour les différentes unités protéiques,  $\alpha$ - et  $\beta$ -globine, sont comprises au sein de molécules d'acide nucléique distinctes.

7 - Procédé selon l'une quelconque des revendications 1 à 6, caractérisé en ce que l'introduction de la ou des molécule(s) d'acide nucléique s'effectue par transformation du génome nucléaire de la cellule végétale.

8 - Procédé selon la revendication 7 caractérisé en ce que la séquence codant pour le composant protéique comprend une ou plusieurs séquence(s) codant pour des signaux d'adressage chloroplastiques, ou pour des signaux d'adressage mitochondriaux.

9 - Procédé selon la revendication 7 caractérisé en ce que la séquence codant pour le composant protéique comprend une ou plusieurs séquence(s) codant pour un peptide signal N-terminal et éventuellement un signal responsable de la rétention de la protéine dans le réticulum endoplasmique, ou un signal d'adressage vacuolaire.

10 - Procédé selon l'une quelconque des revendications 1 à 6 caractérisé en ce que l'introduction de l'acide nucléique s'effectue par la

transformation du génome mitochondrial ou chloroplastique.

11 - Procédé selon l'une quelconque des revendications 1 à 9 caractérisé en ce que l'acide nucléique comprend, outre la ou les séquence(s) codante(s), des séquences régulatrices de transcription reconnues par les cellules végétales.

12 - Procédé selon l'une quelconque des revendications 4 à 11 caractérisé en ce que la ou les séquence(s) codante(s) code(nt) pour une molécule hybride composée d'au moins les parties actives de l' $\alpha$ -globine et de la  $\beta$ -globine.

13 - Procédé selon l'une quelconque des revendications 1 à 12 caractérisé en ce qu'il comprend, entre l'étape de propagation et l'étape de récupération, une étape de détection de protéines héminiques, et en particulier de protéines héminiques dont le noyau porphyrrique est constitué de protoporphyrine IX à fer.

14 - Procédé selon la revendication 13 caractérisé en ce que lors de l'étape de propagation des cellules végétales, de la protoporphyrine IX à fer est additionnée au milieu de culture utilisé pour la croissance des cellules.

15 - Protéine héminique ayant la capacité de fixer réversiblement l'oxygène, caractérisée en ce qu'elle comporte au moins un noyau porphyrrique à fer, d'origine végétale, et un composant protéique comprenant au moins une chaîne polypeptidique, d'origine animale.

16 - Protéine selon la revendication 15 caractérisée en ce que le noyau porphyrrique à fer est la protoporphyrine IX à fer, ou une protoporphyrine se distinguant de la protoporphyrine IX par la nature des

chaines latérales portées par les atomes  $\beta$  des cycles pyrrole.

17 - Protéine selon la revendication 16 caractérisée en ce que le composant protéique comprend au moins une chaîne polypeptidique  $\alpha$ - et/ou  $\beta$ -globine, ou des variantes de celles-ci comportant une ou plusieurs substitution(s), délétion(s) ou insertion(s) d'acides aminés, la protéine héminique étant capable de fixer l'oxygène, de manière réversible.

18 - Protéine selon la revendication 17 caractérisée en ce que la chaîne d' $\alpha$ - ou  $\beta$ -globine, ou les variantes de celles-ci, comprend en outre un signal d'adressage chloroplastique, un signal d'adressage mitochondrial, ou un peptide signal N-terminal éventuellement en combinaison avec un signal responsable de la rétention de la protéine dans le réticulum endoplasmique, ou un signal d'adressage vacuolaire.

19 - Protéine selon la revendication 17, caractérisée en ce que chaque chaîne polypeptidique est dépourvue de Méthionine NH<sub>2</sub>-terminale.

20 - Protéine selon l'une quelconque des revendications 17 à 19, caractérisée en ce que le composant protéique comprend au moins quatre chaînes polypeptidiques d' $\alpha$ - et/ou de  $\beta$ -globine, ou des variantes de celles-ci, chaque chaîne polypeptidique étant liée à un noyau protoporphyrine à fer.

21 - Protéine selon la revendication 20 caractérisée en ce qu'elle comprend 2 chaînes  $\alpha$ -globines et 2 chaînes  $\beta$ -globines, ou des variantes de celles-ci.

22 - Protéine selon l'une quelconque des revendications 15 à 21 caractérisée en ce qu'elle fixe l'oxygène avec une affinité comprise entre 7 et 40 mm Hg, de préférence 15 à 20 mm Hg.

23 - Acide nucléique comportant :

- i) une ou plusieurs séquence(s) codant un composant protéique d'une protéine héminique animale, ladite protéine ayant la capacité de fixer réversiblement l'oxygène, et
- ii) des séquences régulatrices de transcription reconnues par une cellule végétale, comprenant un promoteur et des séquences régulatrices de terminaison, et
- iii) une ou plusieurs séquence(s) codant un signal d'adressage d'origine végétale.

24 - Acide nucléique selon la revendication 23 caractérisé en ce que les séquences régulatrices comprennent un ou plusieurs promoteur(s) d'origine végétale.

25 - Acide nucléique selon la revendication 23 ou 25 caractérisé en ce que les séquences codant le signal d'adressage codent pour un peptide d'adressage mitochondrial ou chloroplastique dit peptide "transit".

26 - Acide nucléique selon la revendication 23 ou 24 caractérisé en ce que les séquences codant le signal d'adressage codent un peptide signal N-terminal d'origine végétale, éventuellement en combinaison avec une séquence codant un signal de rétention endoplasmique ou un signal d'adressage vacuolaire.

27 - Acide nucléique selon l'une quelconque des revendications 23 à 26 caractérisé en ce que la séquence codante code l' $\alpha$ - ou  $\beta$ -globine humaine, ou une variante de celles-ci se différenciant de la séquence naturelle par une ou plusieurs substitution(s), délétion(s) ou remplacement(s) d'acides aminés, ou une partie de l' $\alpha$ - et/ou  $\beta$ -globine humaine.

28 - Acid nucléique selon l'une quelconque des revendications 23 à 27 comportant en outre un ou plusieurs intron(s) de préférence d'origine végétale.

29 - Acide nucléique selon l'une quelconque des revendications 23 à 28 caractérisé en ce que la séquence codant pour le composant protéique est un cDNA.

30 - Vecteur comprenant une ou plusieurs molécule(s) d'acide nucléique selon l'une quelconque des revendications 23 à 29.

31 - Cellules végétales transformées de manière stable par l'acide nucléique selon l'une quelconque des revendications 23 et 29.

32 - Cellules végétales capables de produire un ou plusieurs protéine(s) héminique(s) selon l'une quelconque des revendications 15 à 22.

33. Cellules végétales selon la revendication 32 caractérisées en ce qu'elles comportent de l'acide nucléique comprenant une ou plusieurs séquence(s) codant pour un composant protéique de ladite protéine héminique en association à une ou plusieurs séquence(s) régulatrice(s) de transcription reconnue(s) par la cellule.

34 - Cellules végétales selon l'une quelconque des revendications 31 à 33 caractérisées en ce qu'il s'agit d'une culture de cellules végétales, par exemple en milieu liquide ou immobilisées, ou une culture de racines.

35 - Cellules végétales selon l'une quelconque des revendications 31 à 33 caractérisées en ce qu'il s'agit de cellules faisant partie d'une plante transformée entière.

36 - Plante chimérique ou transgénique capable de produire une ou plusieurs protéine(s) héminique(s), par exemple l'hémoglobine ou un dérivé de celle-ci,

92

caractérisée en ce qu'elle comporte des cellules selon l'une quelconque des revendications 31 à 33.

37 - Graines de plante transgénique selon la revendication 36.

38 - Produit pharmaceutique comprenant une ou plusieurs protéine(s) héminique(s) selon l'une quelconque des revendications 15 à 22 en association avec un excipient acceptable du point de vue physiologique.

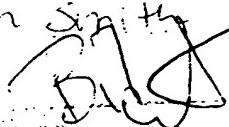
39 - Protéines héminiques selon l'une quelconque des revendications 15 à 22 pour utilisation comme médicament.

40 - Protéine héminique selon la revendication 39 pour une utilisation en thérapie de conditions nécessitant une amélioration du transport d'oxygène dans le sang.

41 - Utilisation d'une protéine héminique selon l'une quelconque des revendications 15 à 22 pour la préparation d'un médicament pour le traitement de conditions nécessitant une amélioration du transport d'oxygène dans le sang.

42 - Utilisation d'une protéine héminique selon l'une quelconque des revendications 15 à 22 dans un produit cosmétique, industriel ou en tant que réactif chimique.

EAI4.22712423US  
16 January 1995

William Smith  


A. CLASSIFICATION OF SUBJECT MATTER  
 IPC 6 C12N15/82 C12N15/12 C07K14/805 C12N5/10 A01H5/00  
 A61K38/42

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
 IPC 6 C12N C07K A01H A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	TRANSFUSION CLINIQUE ET BIOLOGIQUE, vol. 2, no. 6, 1995, pages 441-447, XP000614371 DIERYCK, W., ET AL.: "Recombinant protein expression in plants"  see the whole document ---	1,3,4,7, 11,13, 15-17, 24, 30-36, 38-42
X	WO,A,93 03161 (DONSON JON ;DAWSON WILLIAM O (US); GRANTHAM GEORGE L (US); TURPEN) 18 February 1993 see page 29, line 8 - line 35 *sequence ID nos. 7-10* see page 63, line 15 - line 20	15,18, 23-25, 27,29,30
Y	see page 105 - page 110 ---	1,3-8, 11,13, 31-37
		-/-

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

\* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- \*&\* document member of the same patent family

Date of the actual completion of the international search Date of mailing of the international search report

20 January 1997

23.01.97

Name and mailing address of the ISA  
 European Patent Office, P.B. 5818 Patentlaan 2  
 NL - 2280 HV Rijswijk  
 Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl.  
 Fax (+ 31-70) 340-3016

Authorized officer

Maddox, A

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Description of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO,A,93 25697 (CALIFORNIA INST OF TECHN) 23 December 1993 see claims 1,17	23,24, 27,30-37 1,3-8, 11,13, 31-37
X	---	17, 20-22, 39-42
X	WO,A,93 08831 (STROHTECH INC) 13 May 1993  see claims 71-74	27,30-34
X	---	
X	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 90, WASHINGTON US, pages 11811-11815, XP002023062 OHME-TAKAGI, M., ET AL.: "The effect of sequences with high AU content on mRNA stability in tobacco" see the whole document	27,30-34
X	---	
X	BIOTECHNOLOGY, vol. 9, no. 1, pages 57-61, XP002023063 WAGENBACH, M., ET AL.: "Synthesis of wild type and mutant human hemoglobins in <i>Saccharomyces cerevisiae</i> " see the whole document	17,19-22
A	---	
A	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 88, no. 16, 15 August 1991, pages 7041-7045, XP000272765 KAZUKI SAITO ET AL: "INTEGRATION AND EXPRESSION OF A RABBIT LIVER CYTOCHROME P-450 GENE IN TRANSGENIC NICOTIANA TABACUM" see the whole document	1,2,7, 11,13, 23,24, 29-37
A	---	
A	WO,A,91 06320 (SCRIPPS CLINIC RES) 16 May 1991 see page 42, line 29 - line 35	1-42
A	---	
A	EP,A,0 273 889 (GENELUX AB) 6 July 1988 see the whole document	1-42
	-----	

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A-9303161	18-02-93	US-A-	5529909	25-06-96
		US-A-	5589367	31-12-96
		US-A-	5316931	31-05-94
		AU-A-	3351193	02-03-93
		CA-A-	2114636	18-02-93
		EP-A-	0596979	18-05-94
		JP-T-	7503361	13-04-95
-----	-----	-----	-----	-----
WO-A-9325697	23-12-93	NONE		-----
-----	-----	-----	-----	-----
WO-A-9308831	13-05-93	NONE		-----
-----	-----	-----	-----	-----
WO-A-9106320	16-05-91	US-A-	5202422	13-04-93
		AU-A-	6753290	31-05-91
		CA-A-	2071502	28-04-91
		EP-A-	0497904	12-08-92
		JP-T-	5504333	08-07-93
-----	-----	-----	-----	-----
EP-A-0273889	06-07-88	JP-A-	1013993	18-01-89
-----	-----	-----	-----	-----

1/11

### Structure de l'hème

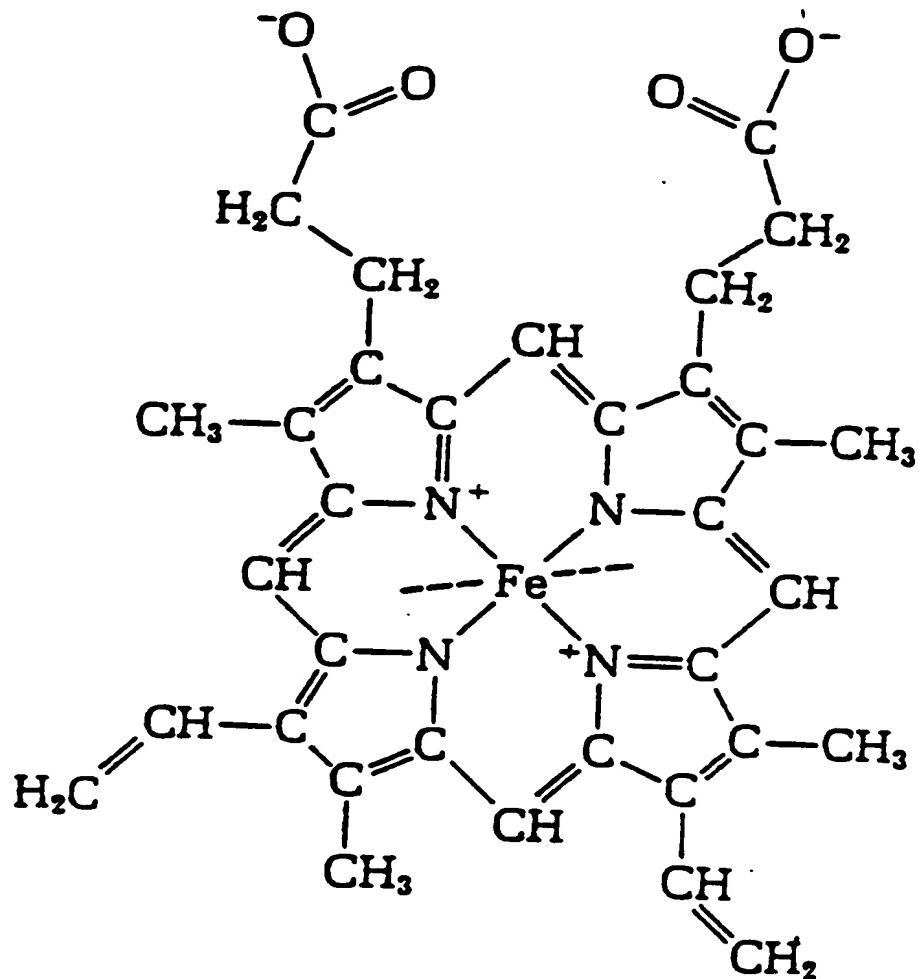


FIGURE 1

2/11

Séquence alpha globine:

1	GTG	TCT	CCT	GCC	GAC	AAG	ACC	AAC	GTC	AAG	GCC	GCC	45	
	Val	Leu	Ser	Pro	Ala	Asp	Lys	Thr	Asn	Val	Lys	Ala	Trp	Gly
46	AAG	GTT	GCG	GCG	CAC	GCT	GCC	GAG	TAT	GGT	GCG	GAG	GAG	90
	Lys	Val	Gly	Ala	His	Ala	Gly	Glu	Tyr	Gly	Ala	Glu	Ala	Glu
91	AGG	ATG	TTC	CTG	TCC	TTC	CCC	ACC	ACC	AAG	ACC	TAC	TTC	135
	Arg	Mel	Phe	Leu	Ser	Phe	Pro	Thr	Thr	Lys	Thr	Tyr	Phe	Pro
136	TTC	GAC	CTG	AGC	CAC	GCC	TCT	GCC	CAG	GTT	AAG	GCC	CAC	180
	Phe	Asp	Leu	Ser	His	Gly	Ser	Ala	Gln	Val	Lys	Gly	His	Gly
181	AAG	GTG	GCC	GAC	GCG	CTG	ACC	AAC	GCC	GTG	GCG	CAC	GAC	225
	Lys	Val	Ala	Asp	Ala	Leu	Thr	Asn	Ala	Val	Ala	His	Val	Asp
226	ATG	CCC	AAC	GCG	CTG	TCC	GCC	CTG	AGC	GAC	CTG	CAC	GCG	270
	Mel	Pro	Asn	Ala	Leu	Ser	Ala	Leu	Leu	Ser	Asp	Leu	His	Ala
271	CTT	CGG	GTG	GAC	CCG	GTC	AAC	TTC	AAG	CTC	CTA	AGC	CAC	315
	Leu	Arg	Val	Asp	Pro	Val	Asn	Phe	Lys	Leu	Leu	Ser	His	Cys
316	CTG	GTG	ACC	CTG	GCC	GCC	CTC	CCC	GCC	GAG	TTC	ACC	CCT	360
	Leu	Val	Thr	Leu	Ala	Ala	His	Leu	Pro	Ala	Glu	Phe	Thr	Pro
361	GTG	CAC	GCC	TCC	CTG	GAC	AAG	TTC	CTG	GCT	TCT	GTG	AGC	405
	Val	His	Ala	Ser	Leu	Asp	Lys	Phe	Leu	Ala	Ser	Val	Ser	Thr
406	CTG	ACC	TCC	AAA	TAC	CGT								
	Leu	Thr	Ser	Lys	Tyr	Arg								

FIGURE 2

3/11

## Séquence bêta globine:

1	GTG	CAC	CTG	ACT	CCT	GAG	AAG	TCT	GCC	GTT	ACT	GCC	CTG	TGG	4 5	
	Val	His	Leu	Thr	Pro	Glu	Glu	Lys	Ser	Ala	Val	Thr	Ala	Leu	Trp	
46	GGC	AAG	GTG	AAC	GTG	GAT	GAA	GTT	GGT	GAG	GCC	CTG	GGC	AGG	9 0	
	Gly	Lys	Val	Asn	Val	Asp	Glu	Val	Gly	Gly	Glu	Ala	Leu	Gly	Arg	
91	CTG	CTG	GTT	GTC	TAC	CCT	TGG	ACC	CAG	AGG	TTC	TTT	GAG	TCC	135	
	Leu	Leu	Val	Tyr	Pro	Trp	Thr	Gln	Arg	Phe	Phe	Glu	Ser	Phe		
136	GCG	GAT	CTG	TCC	ACT	CCT	GAT	GCT	GTT	ATG	GCC	AAC	CCT	AAG	GTG	180
	Gly	Asp	Leu	Ser	Thr	Pro	Asp	Ala	Val	Met	Gly	Asn	Pro	Lys	Val	
181	AAG	GCT	CAT	GGC	AAG	AAA	GTG	CTC	GGT	GCC	TTT	AGT	GAT	GGC	CTG	225
	Lys	Ala	His	Gly	Lys	Lys	Val	Leu	Gly	Ala	Phe	Ser	Asp	Gly	Leu	
226	GCT	CAC	CTG	GAC	AAC	CTC	AAG	GCG	ACC	TTT	GCC	ACA	CTG	AGT	GAG	270
	Ala	His	Leu	Asp	Asn	Leu	Lys	Gly	Thr	Phe	Ala	Thr	Leu	Ser	Glu	
271	CTG	CAC	TGT	GAC	AAG	CTG	CAC	GTG	GAT	CCT	GAG	AAC	TTC	AGG	CTC	315
	Leu	His	Cys	Asp	Lys	Leu	His	Val	Asp	Pro	Glu	Asn	Phe	Arg	Leu	
316	CTG	GCG	AAC	GTG	CTG	GTC	TGT	GTG	CTG	GCC	CAT	CAC	TTT	GGC	AAA	360
	Leu	Gly	Asn	Val	Leu	Val	Cys	Vai	Leu	Ala	His	His	Phe	Gly	Lys	
361	GAA	TTC	ACC	CCA	CCA	GTG	CAG	GCT	GCC	TAT	CAG	AAA	GTG	GTG	GCT	405
	Glu	Phe	Thr	Pro	Pro	Val	Gln	Ala	Ala	Tyr	Gln	Lys	Val	Val	Ala	
406	GGT	GTG	GCT	AAT	GCC	CTA	GCC	CAC	AAG	TAT	CAC					
	Gly	Val	Ala	Asn	Ala	Leu	Ala	His	Lys	Tyr	His					

FIGURE 3

4/11

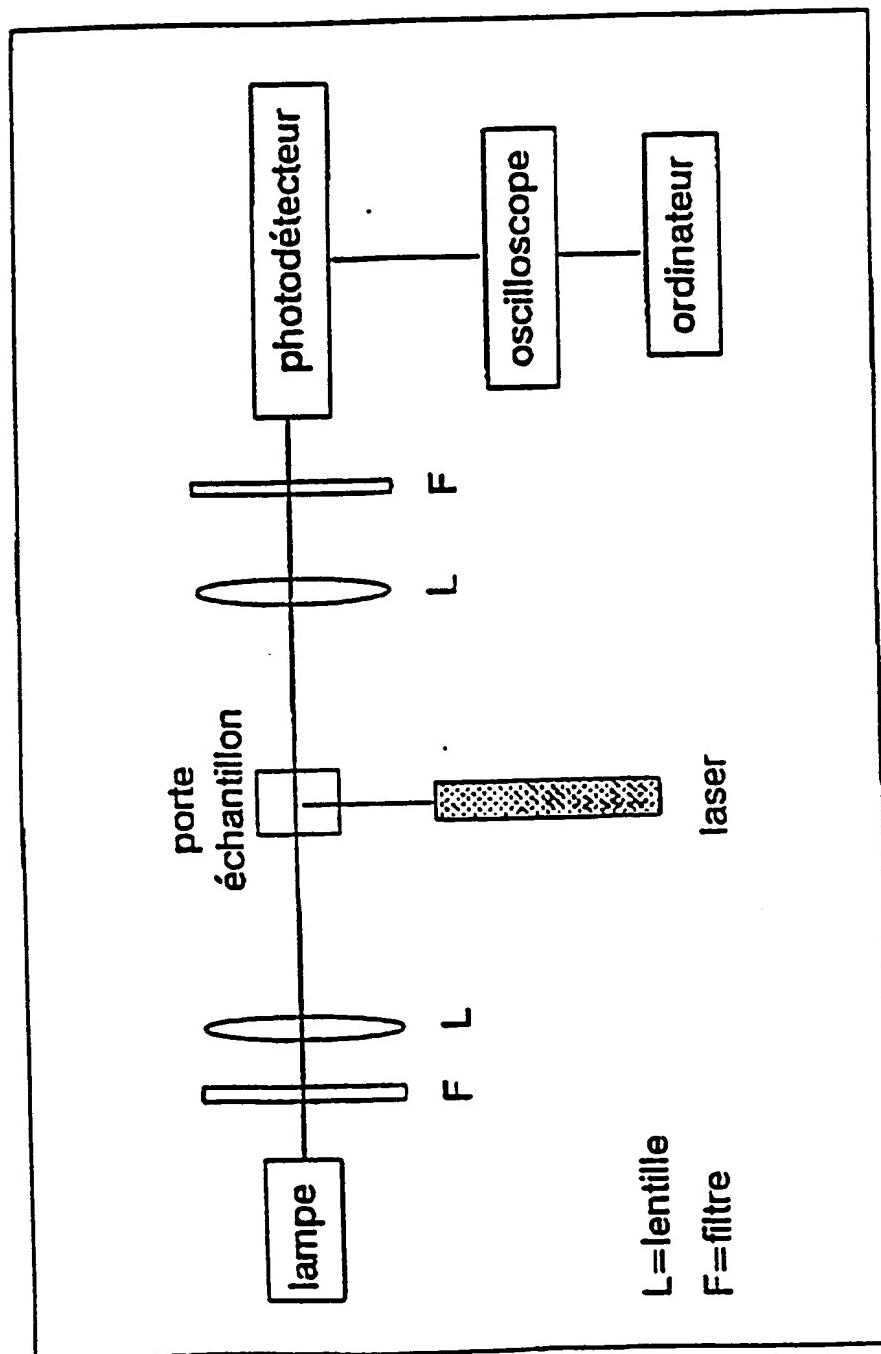


FIGURE 4

5/11

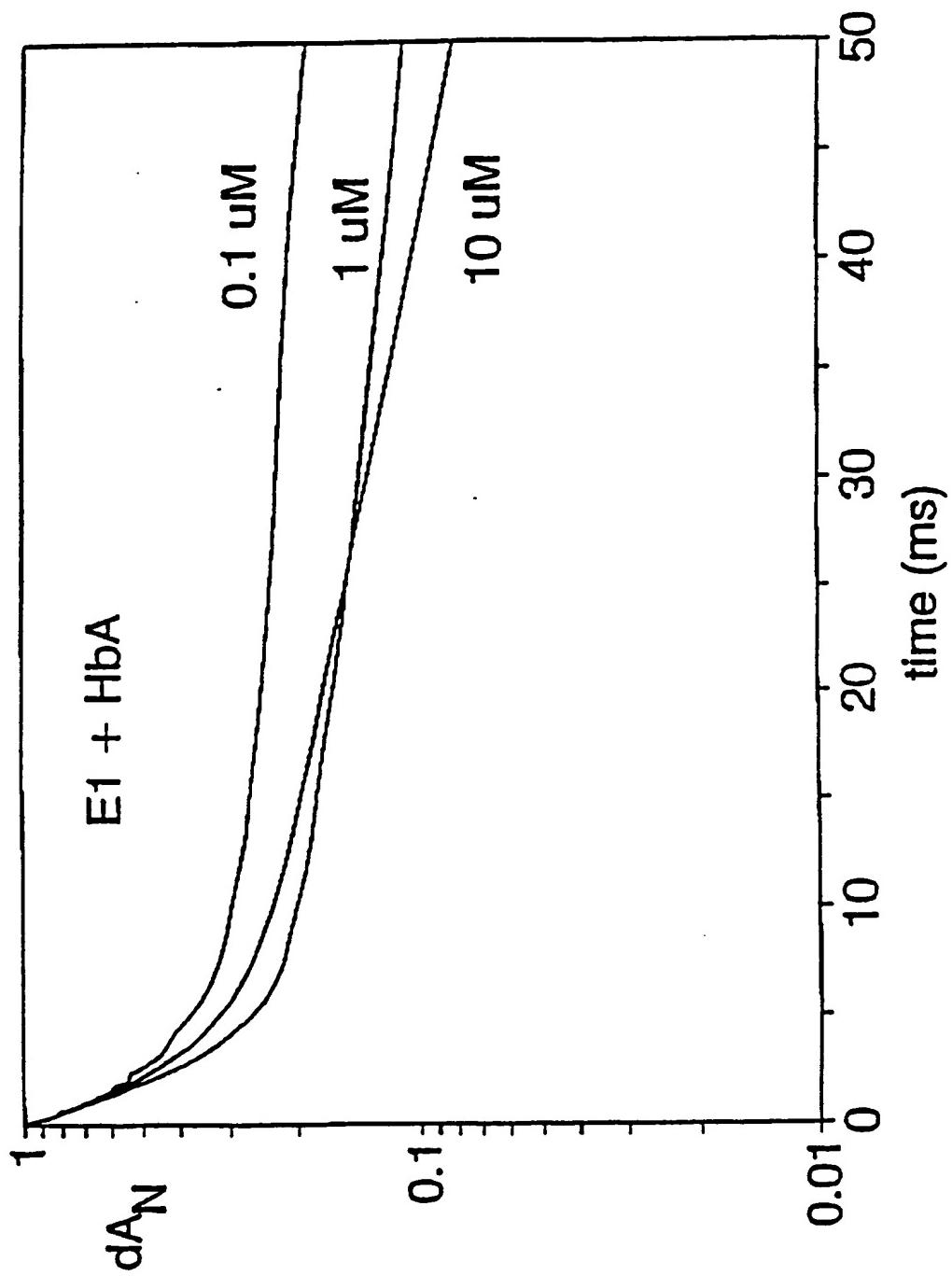


FIGURE 5

6/11

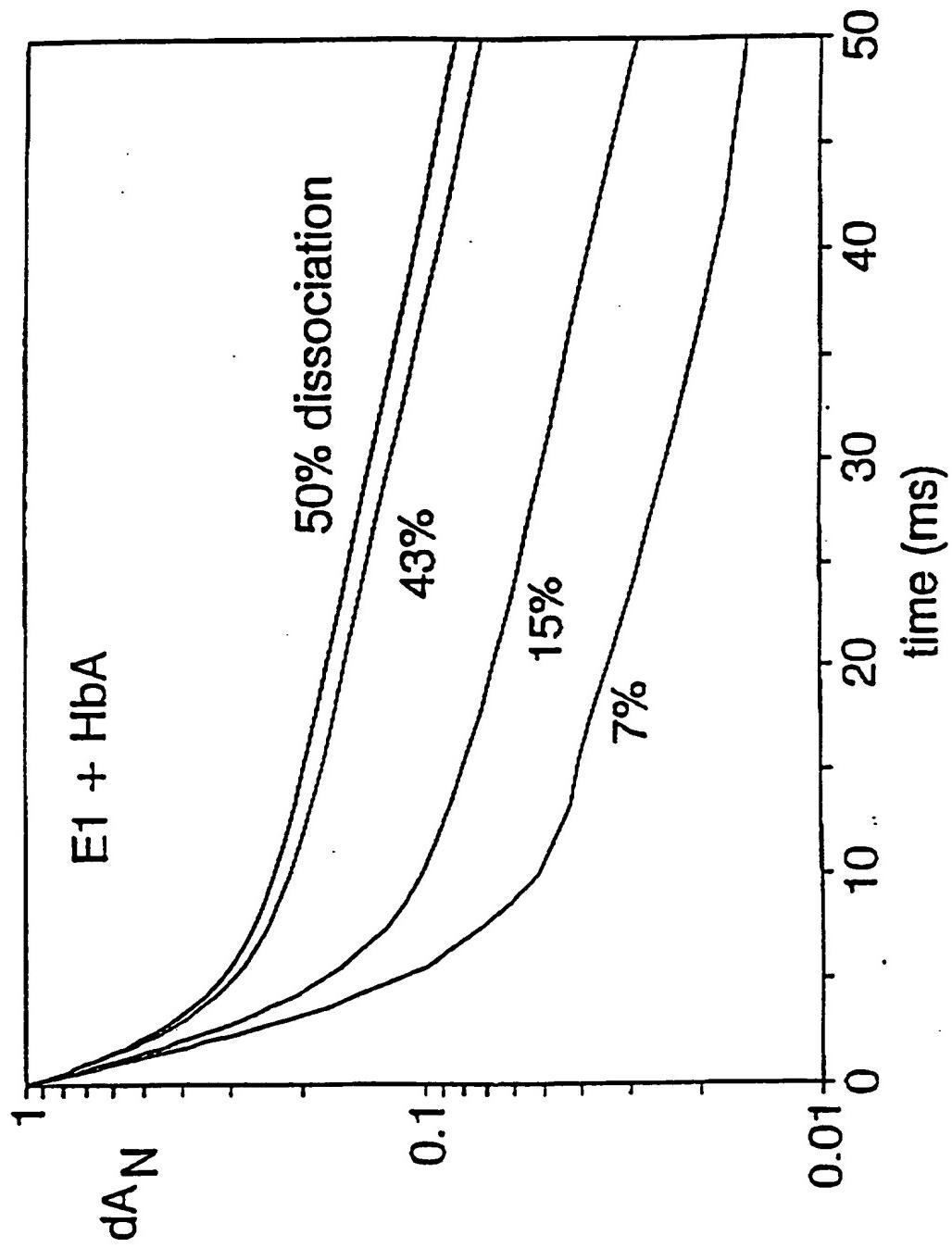


FIGURE 6

7/11

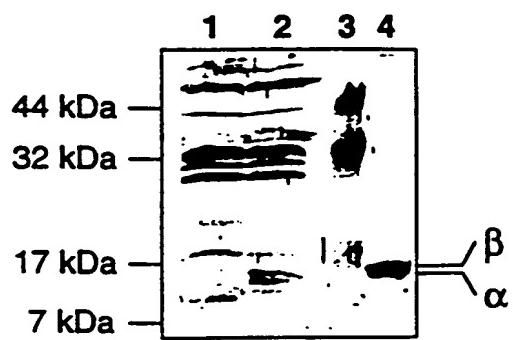


FIGURE 7

8/11

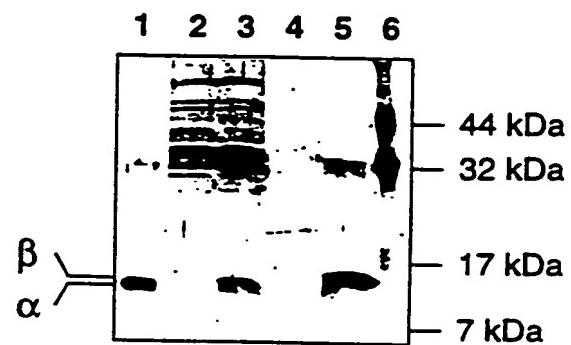


FIGURE 8

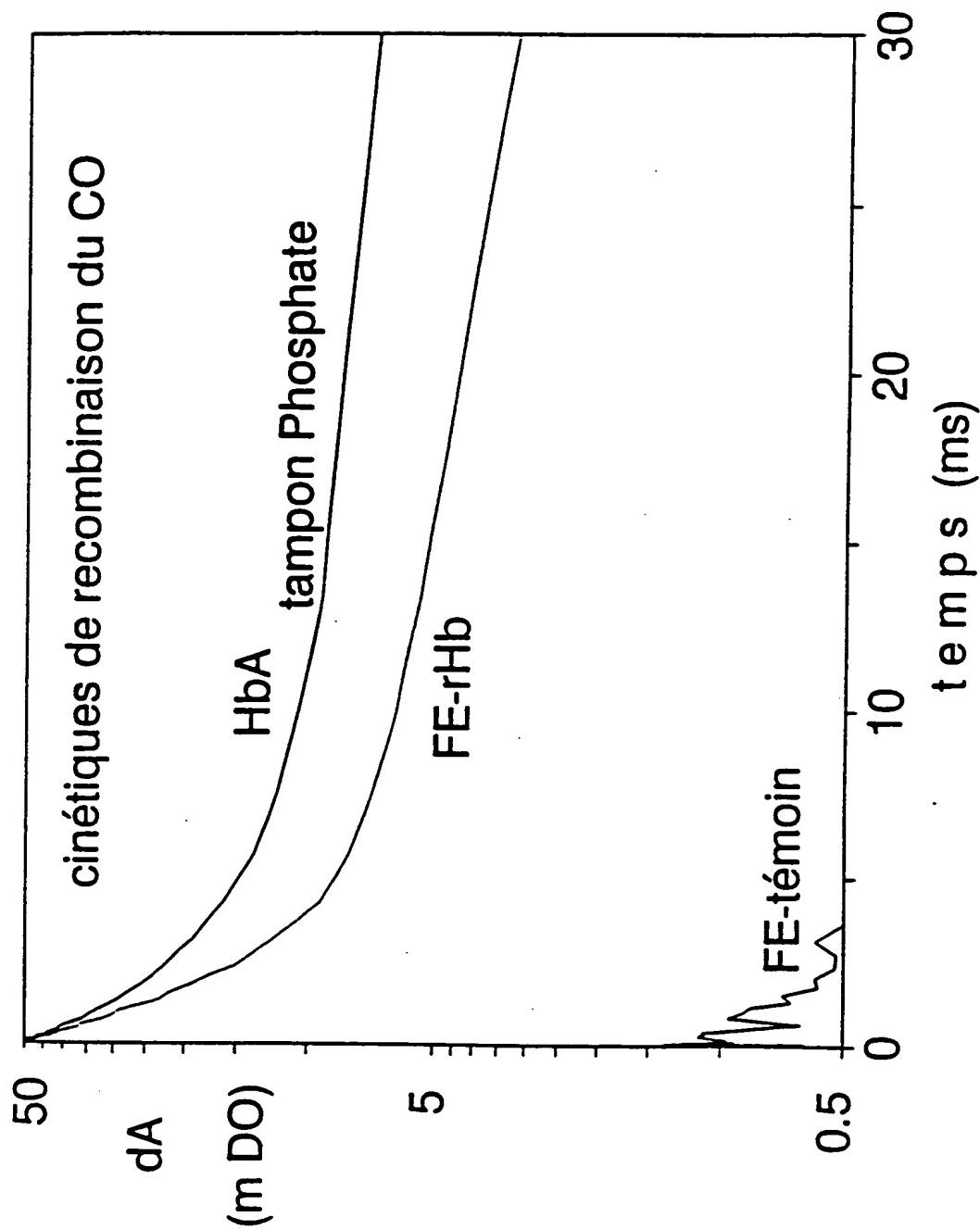


FIGURE 9

10/11

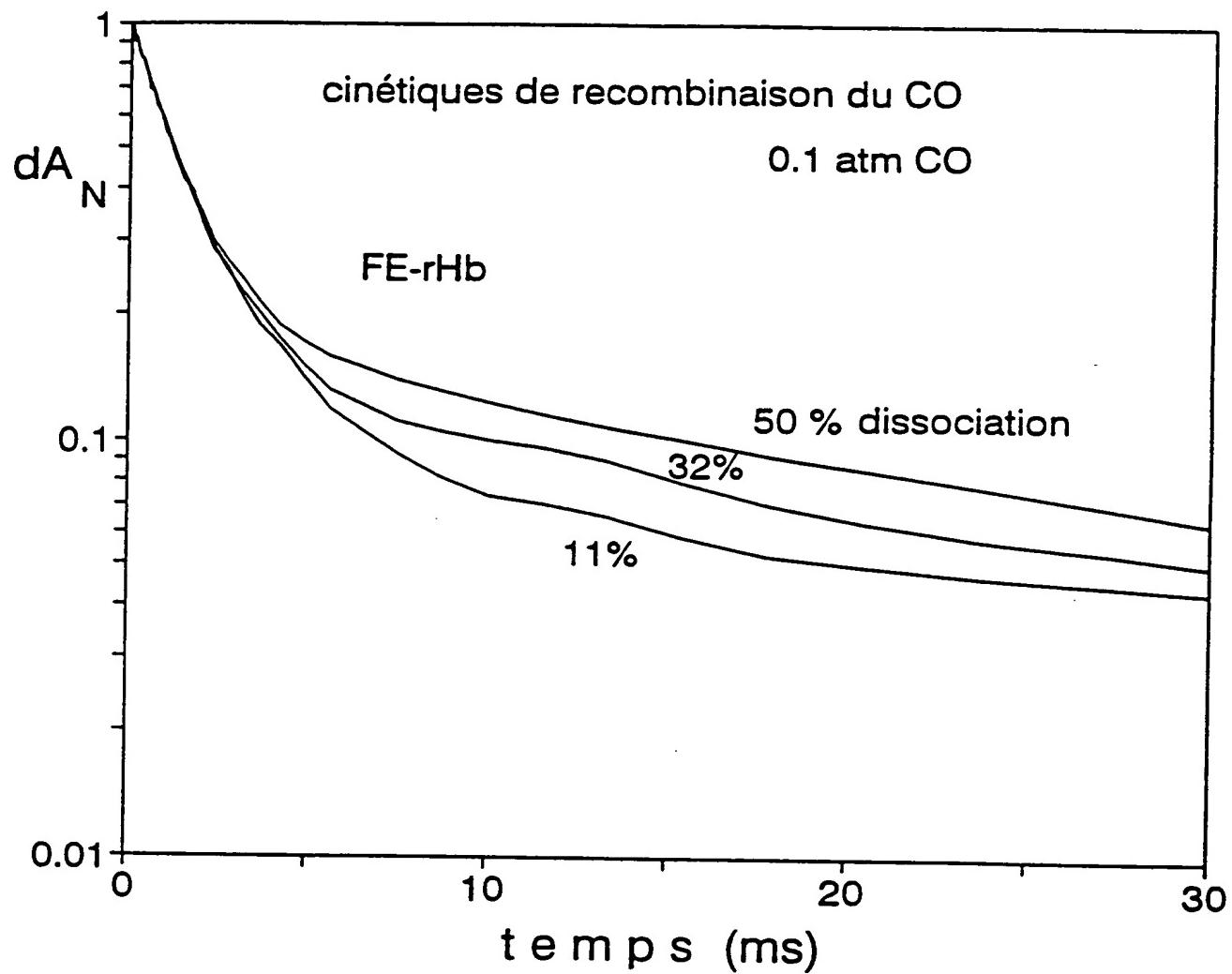


FIGURE 10

11/11

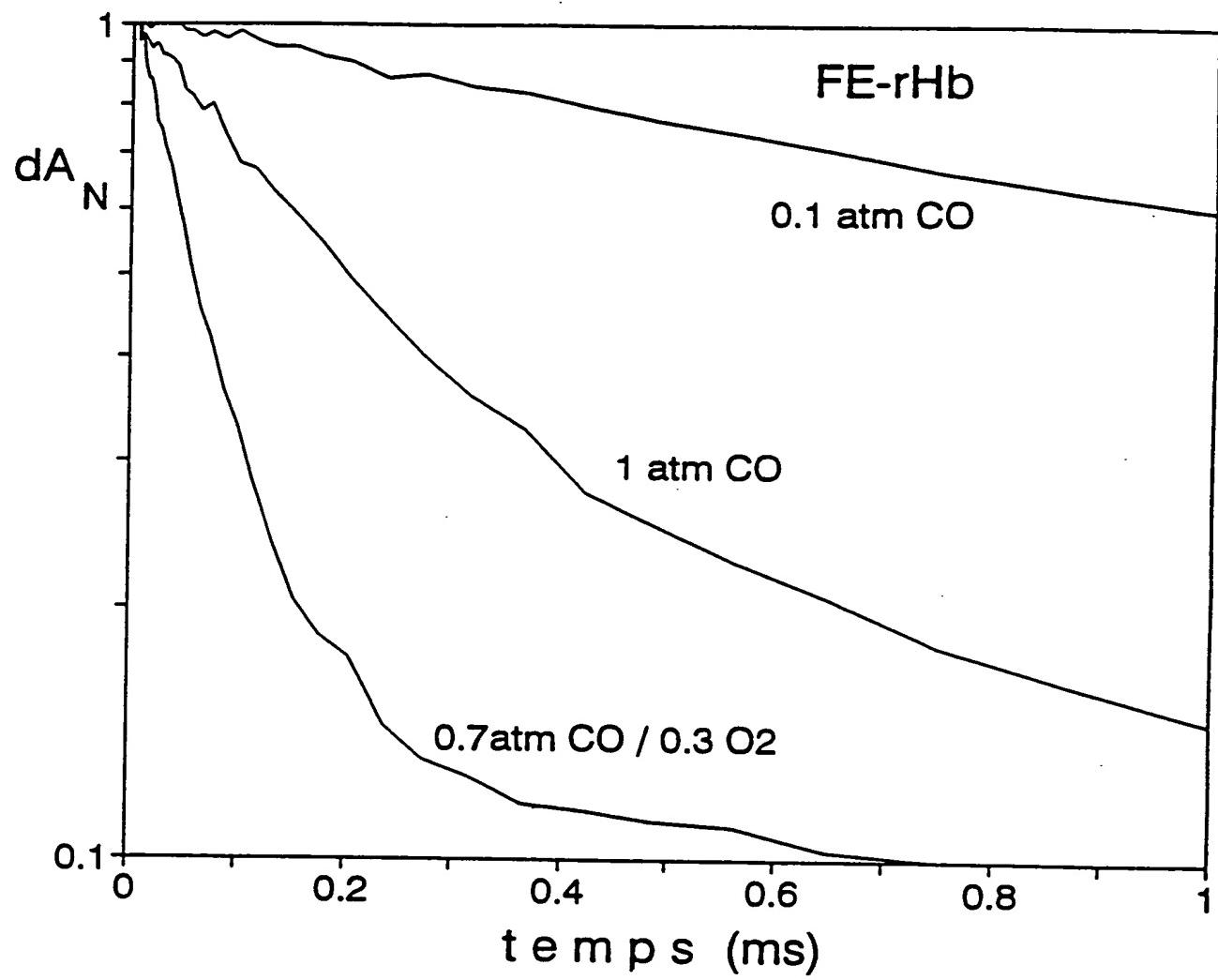


FIGURE 11

WO 97/04115

PCT/FR96/01123

METHOD FOR PRODUCING HEMIN PROTEINS USING PLANT CELLS,  
RESULTING PROTEINS AND PRODUCTS CONTAINING SAME

*Jms  
JAI* >

The invention relates to a method for producing hemin proteins using plant cells, and in particular the hemin proteins capable of reversibly binding oxygen, for example hemoglobin and its derivatives, and myoglobin. It relates, in addition, to the proteins obtained using the method. The invention also relates to the genetically transformed cells and plants capable of producing these proteins, and to the nucleic acid constructs involved in the transformation. In addition, the invention relates to products, for example pharmaceutical or cosmetic products, containing these hemin proteins.

Hemin proteins are complex molecules composed of one or more polypeptide chain(s) in association with one or more ferroporphyrin nucleus or nuclei. These nuclei are composed of four pyrrole rings, juxtaposed in a closed structure and linked by methene bridges, and containing an iron atom at the center of the molecule. Hemin proteins differ from one another in the nature and the number of the polypeptide chains and in the nature of the side chains carried by the eight  $\beta$  atoms of the pyrrole rings. An example of a ferroporphyrin nucleus is iron-containing protoporphyrin IX, also known by the name "protoheme" or simply "heme" (Figure 1).

The hemin protein family comprises numerous substances which are important from the biological point of view in animals and in plants, particularly hemoglobin, myoglobin, cytochromes, peroxidases and catalases.

Hemoglobin is the main constituent of the red blood cells. Its essential function is to bind, transport and deliver the quantity of oxygen necessary for normal tissue function.

The hemoglobin molecule is composed of two types of globin chains or subunits, called  $\alpha$  and  $\beta$  (of 141 and 146 amino acids respectively), and linked in pairs to form a  $\alpha_2\beta_2$  tetramer. Each of these subunits contains, solidly attached in a hydrophobic sac, a heme molecule (that is to say protoporphyrin IX) containing, at the center, a divalent iron atom ( $Fe^{2+}$ ) to which a molecule of oxygen reversibly binds. Each tetrameric hemoglobin molecule therefore contains 4 iron atoms and 4 oxygen

*JfA*

molecules which it binds during its passage through the lungs. The molecular mass of the tetramer is 64,650 D. In man, the  $\alpha$  and  $\beta$  chains are synthesized from two types of genes situated on chromosomes 16 and 11 respectively.

The term beta, or "nonalpha", type chains covers not only the beta chains, but also the chains called epsilon, gamma or delta.

Normally, in adults, more than 95% of the hemoglobin consists of alpha<sub>2</sub> beta<sub>2</sub> tetramer, that is to say the association of two heterologous alpha-beta dimers, associated with the catalytic complex, heme. 2% to 3% of a hemoglobin consisting of alpha<sub>2</sub> delta<sub>2</sub> tetramers, and traces of fetal hemoglobin alpha<sub>2</sub> gamma<sub>2</sub> exist.

The tetrameric human hemoglobin molecule exists in two quaternary forms or structures depending on whether oxygen is bound or not to the iron atoms. In the presence of oxygen, hemoglobin is said to be in an R (for relaxed) state and its affinity for oxygen is high. In the absence of oxygen, hemoglobin is said to be in a T (for tense) state and its affinity for oxygen is 100 times lower (Perutz, 1970). The resultant affinity is linked to the equilibrium between the concentrations of R and T forms. The higher the concentration of hemoglobin in the T form at any level of oxygenation, the lower this affinity. The affinity of hemoglobin for oxygen is regulated by the cofactor - 2,3-diphosphoglycerate (DPG), a small molecule derived from the metabolism of glucose and which binds to the  $\beta$  chains of tetrameric hemoglobin, reducing its affinity for oxygen.

The increase in the risks of infection by products derived from human blood (hepatitis, HIV) makes the development of an artificial oxygen carrier as substitute for blood transfusion necessary.

Techniques using recombinant DNAs have been proposed for producing the protein chains of globin.

The aim of the first techniques developed was essentially to cause the alpha and beta chains to be synthesized in E. coli separately (Nagaï and Thogen-Sen, 1987), involving cumbersome methods for separate expression of each of the chains. These methods could hardly be exploited on an industrial scale.

More recently, the expression of soluble and functional recombinant hemoglobin has been developed in E. coli (Hoffman et al., 1990, P.N.A.S., 87, 8521-

8525) and Saccharomyces cerevisiae (Wagenbach et al., 1991, Biotechnology, 9, 57-61). Each of these systems has advantages and disadvantages. Indeed, the highest expression levels are obtained in E. coli which has, nevertheless, the disadvantage of producing endotoxins and of not cleaving the NH<sub>2</sub> terminal methionines contrary to Saccharomyces cerevisiae. In the yeast, the yields of synthesis of hemoglobin are low (3 to 5%), compared with the 10-15% obtained in E. coli. This currently limits the use of yeast in the context of an industrial development plan.

The use of animal cells in culture or of transgenic animals as hosts for production has also been achieved (Swanson et al., Bio/Technology, May 1992, 10, page 55). It appears that these techniques cannot currently be exploited because of low expression levels and the risks of contaminations by viruses and by prions.

The technical problem which the present invention purposed to solve is to produce hemin proteins, and in particular hemoglobin and its derivatives, in a large quantity at low costs, without the risk of viral or subviral contaminations. The inventors have provided a solution to this problem by using plant cells as host for the transformation and the production.

Various teams have already taken an interest in the production of mammalian recombinant proteins in plant cells or in transgenic plants. For example, the specific expression, in rapeseed, of leu-enkephalin has been obtained with expression levels of about 0.1% (Vanderkerckhove et al., Biotechnology, 1989, 7, 929-932).

In 1990, Sijmons et al., (Biotechnology, 1990, 8, 217-221) transferred the gene for human serum albumin into tobacco and potato cells. Regardless of the origin of the signal peptides (human or plant), human serum albumin levels of the order of 0.02% of the total proteins were obtained in the potato leaves, stems and tubers.

Other mammalian recombinant proteins have also been produced in plants: hepatitis B surface antigen (Mason et al., P.N.A.S., 1992, 89, 11745-11749); human interferon (Edelbaum J. of Interferon Res., 1992, 12, 449-453); a mouse antibody to Streptococcus mutans, an agent for dental caries (Hiatt and Ma, FEBS, 1992, 307, 71-75); an anti-Herpes antibody (Russel, 1994) and hirudin (Moloney, 1994).

All these research studies show that the production of mammalian recombinant proteins in plant cells is possible and that the mechanisms of protein synthesis from the DNA sequences are similar in animal cells and plant cells.

On the other hand, little information is available on the subject of the iron-containing porphyrins in plants, particularly on their structures, their synthesis pathways and the assembly of the porphyrin nuclei and the protein chains to form the hemin proteins. The production of recombinant molecules having the capacity to reversibly bind oxygen, and requiring the assembly, in the cell, of heterologous proteins and of endogenous plant porphyrins has never been described.

The invention relates to a method for producing recombinant hemin proteins using plant cells. According to the method of the invention, the plant cell is genetically modified so as to be able to express the protein component of a hemin protein. The porphyrin nucleus is produced by the cell endogenously, the assembling of the protein and porphyrin components taking place spontaneously by virtue of their high affinity for each other.

More particularly, the invention relates to a method for producing hemin proteins comprising the following steps:

- i) introducing, into plant cells, one or more nucleic acid molecule(s) each of which comprises at least one sequence encoding a protein component of a hemin protein of animal origin or a variant or a portion of this protein component, and optionally a sequence encoding a selection agent;
  - ii) selecting the cells which have integrated the nucleic acid encoding the protein component;
  - iii) propagating the transformed cells, either in culture or by regenerating whole transgenic or chimeric plants;
  - iv) recovering, and optionally purifying, a hemin protein comprising a complex of the protein or proteins encoded by the abovementioned nucleic acid with at least one iron-containing porphyrin nucleus, or a plurality of these complexes.

The invention preferably relates to a method for producing hemin proteins comprising the following steps:

- i) introducing, into plant cells, one or more nucleic acid molecule(s) each of which comprises at least one sequence encoding a protein component of a hemin protein of animal origin preferably capable of reversibly binding oxygen or a variant or a portion of this protein component, and optionally a sequence encoding a selection agent;
- ii) selecting the cells which contain the nucleic acid encoding the protein component of the hemin protein;
- iii) optionally, propagating the transformed cells, either in culture or by regenerating whole transgenic or chimeric plants;
- iv) recovering, and optionally purifying, a hemin protein comprising a complex consisting of the protein or proteins encoded by the abovementioned nucleic acid and at least one iron-containing porphyrin nucleus, or a plurality of these complexes.

In the context of the present invention, the term "hemin protein" means any protein having an iron-containing porphyrin nucleus as prosthetic group, and in particular protoporphyrin IX as exists in human myoglobin and hemoglobin (Figure 1). The porphyrin nucleus may also be derivatives of heme from those of human heme. The side chains are preferably hydrophobic.

The hemin proteins of the invention include in particular the hemin proteins having, as main function, the reversible binding of oxygen, that is to say myoglobin and hemoglobin, as well as the cytochromes whose role is to transport electrons. The derivatives of these proteins conserving these functions are also included in the invention.

According to a preferred variant, the hemin protein of the invention is hemoglobin or a hemoglobin-type protein. In the context of this invention, the term "hemoglobin-type protein" includes all the hemin proteins having at the same time:

- i) one or more  $\alpha$ - and/or  $\beta$ -globin chain(s) or variants of these polypeptides, and
- ii) one or more molecules of iron-containing protoporphyrin IX, or of protoporphyrins differing from protoporphyrin IX in the nature of the side chains,
- iii) having a capacity to reversibly bind oxygen, preferably with an affinity of between 10 and 50 mm Hg at 37°C, pH7.4. More particularly, the affinity is between 20 and 30 mm Hg, by way of example, the  $P_{50}$  of total blood at pH 7.2 is of  $26 \pm 2$  mm Hg.

In the text which follows, the term "hemoglobin-type molecule" will be used synonymously with the term "hemoglobin derivative".

In this context, a "variant" of a protein component, and particularly of  $\alpha$ - or  $\beta$ -globin, means an amino acid sequence which distinguishes itself in relation to the natural sequence by one or more amino acid substitution(s), deletion(s) or insertion(s). In general, the variant exhibits at least 90%, and preferably at least 95%, homology or identity with the natural sequence. In the context of the present invention, the percentage homology between two amino acid sequences is calculated as being the number of identical amino acids plus the number of similar amino acids in the alignment of the two sequences, divided by the length of the sequences between two given positions. If, between the two given positions, the two sequences do not have the same length, the percentage homology is the number of identical and similar amino acids, divided by the length of the longest sequence. The amino acids considered to be "similar" are known in the art, see for example R.F. Feng, M.S. Jobson and R.F. Doolittle; J. Mol. Evol.; 1985; 21; 112-115. They are normally considered to be those which, within a permutation matrix, have a positive coefficient of substitution.

The term "variant" also includes fragments of polypeptide chains, for example of  $\alpha$ - or  $\beta$ -globin, normally having a length of at least 90% of the parent molecule. The variants can also be made longer than the parent molecule by adding nonfunctional sequences. Preferably, the variants conserve the biological and immunological properties of the parent molecule.

The first stage of the method of the invention consists in introducing, into plant cells, one or more nucleic acid molecule(s) comprising at least one sequence encoding a protein component of a mammalian hemin protein, or a variant of this component.

When the hemin protein is a single-chain protein, for example myoglobin or cytochrome, the nucleic acid introduced into the plant cells normally comprises a copy of the sequence encoding this protein.

On the other hand, when it is an oligomeric or a multimeric protein, such as hemoglobin or hemoglobin-type molecules, the sequences encoding the various protein units are introduced into the plant cell, either within the same nucleic acid

molecule, or within separate nucleic acid molecules. Preferably, for the production of hemoglobin and its derivatives, the sequences encoding  $\alpha$ - and  $\beta$ -globin, or their variants, are within the same vector, called co-expression vector. The vector may comprise one or more copy(ies) of each encoding sequence.

Alternatively, the sequences encoding  $\alpha$ - and  $\beta$ -globin, or their variants, may be present on separate nucleic acid molecules. According to this variant, the two molecules may be introduced into the same plant cell, provided that an appropriate selection system is available. Another technique consists in introducing one of the molecules into a first plant cell, and the other into a second plant cell. Each of the transformed cells is then regenerated into a whole plant, it then being possible for the plants thus obtained to be crossed in order to give a progeny capable of producing both the  $\alpha$  and  $\beta$  chains. This approach can be used to optimize the yield of hemoglobin.

The nucleic acid molecules introduced into the plant cell during the first stage of the method are also part of the invention. Generally, these nucleic acids comprise:

- i) one or more sequence(s) encoding a protein component of an animal hemin protein, and
- ii) one or more sequence(s) encoding a targeting signal of plant origin, and/or sequences for regulation of transcription which are recognized by a plant cell.

More particularly, the nucleic acid of the invention comprises:

- i) one or more sequence(s) encoding a protein component of an animal hemin protein, the said protein having the capacity to reversibly bind oxygen, and
- ii) sequences for regulation of transcription which are recognized by a plant cell, comprising a promoter and sequences for regulation of termination, and
- iii) one or more sequence(s) encoding a targeting signal of plant origin.

Preferably, the sequences encoding the protein component encode animal  $\alpha$ - or  $\beta$ -globin, for example of human or bovine origin, or the variants thereof. In this manner, the properties of the molecule, and in particular the affinity for oxygen and the stability, can be optimized.

Among these modifications, it is possible, for example, to introduce into one or into both of the  $\alpha$ - and  $\beta$ -globin chains, by site-directed mutagenesis, one or two

sequence difference(s) in order to reduce the affinity for oxygen. These mutations may be chosen from examples of natural mutations (see Table I), or from the mutations indicated by examination of the three-dimensional model of natural hemoglobin A.

Table I: Some mutated human hemoglobins

(Int. Hemoglobin Center, 1995)

Abnormal hemoglobin	Normal residues and positions	Replacement
$\alpha$ chain		
I	16 Lys	Glu
G <sub>Honolulu</sub>	30 Glu	Gln
Norfolk	57 Gly	Asp
M <sub>Boston</sub>	58 His	Tyr
G <sub>Philadelphia</sub>	68 Asn	Lys
O <sub>Indonesia</sub>	116 Glu	Lys
$\beta$ chain		
C	6 Glu	Lys
S	6 Glu	Val
G <sub>San José</sub>	7 Glu	Gly
E	26 Glu	Lys
M <sub>Saskatoon</sub>	63 His	Tyr
Zurich	63 His	Arg
M <sub>Milwaukee</sub>	67 Val	Glu
D <sub>Punjab</sub>	121 Glu	Gln
Mequon	41 Phe	Tyr
Providence	82 Lys	Asp

In a very advantageous manner, the mutants whose functional properties correspond to the physiological conditions for oxygen transport will be used: reversible

binding, cooperativity and low speed of autooxidation. Among the mutants, there will be preferably used the double mutants  $\alpha_2\beta_2F41Y,K82D$  (that is to say a mutant whose  $\beta$  chain comprises the following modifications: Phe-41 is replaced by Tyr, and Lys-82 is replaced by Asp) or  $\alpha_2\beta_2F41Y,K66T$  (that is to say a mutant whose  $\beta$  chain comprises the following modifications: Phe-41 is replaced by Tyr, and Lys-66 is replaced by Thr) which correspond to these functional characteristics.

The modifications of the  $\alpha$  and  $\beta$  chains may also be carried out in order to stabilize the molecule, that is to say to avoid the dissociation of the tetramer into small-sized dimers which are rapidly filtered by the kidneys and which limit the intravascular life of hemoglobin. Covalent bridging, with the aid of phosphate or diaspirin, has been demonstrated as being an effective technique for stabilizing the tetramer (Benesch and Kwong, 1994). The same result can be obtained through modifications of the amino acid chain. The  $\alpha$  subunits are produced in an alpha-alpha dimeric form linked by a glycyl residue. In this form, they conserve their capacity to correctly assemble onto the beta partner subunits and onto heme in order to form a soluble hemoglobin. This hemoglobin can no longer dissociate into dimers because the tetrameric structure is stabilized by a covalent bond (peptide bond) between the alpha-beta dimers. This technique makes it possible to increase the intravascular half-life of the molecule.

Among the variants, it is also possible to use a hybrid protein composed of a portion of the alpha chain and a portion of the beta chain.

According to a preferred variant of the invention, the nucleic acid comprises, in addition to the sequences encoding  $\alpha$ - or  $\beta$ -globin, sequences encoding targeting signals. Preferably, these signals are chloroplast or mitochondrial targeting signals. The expression and/or accumulation of the recombinant proteins in these organelles is particularly preferred because of the availability of endogenous iron-containing porphyrins which are found here. The yield of hemin proteins is therefore increased. In addition, the targeting of the proteins toward the chloroplasts and the mitochondria avoids glycosylation of the protein, which may be advantageous since the natural hemoglobin molecule is not glycosylated.

As an example of chloroplast targeting signals, there may be mentioned the sequence encoding the transit peptide of the precursor of the small subunit of ribulose 1,5-diphosphate carboxylase of Pisum sativum (see examples). As mitochondrial targeting signals, there may be mentioned the sequence encoding the transit peptide of the precursor of the beta subunit of mitochondrial ATP-aseF1 of Nicotiana plumbaginifolia (see examples).

These transit peptides, as well as the N-terminal methionine, are normally cleaved in the chloroplasts or the mitochondria. The expression of the proteins in the plastids therefore also has the advantage of producing a molecule lacking N-terminal methionine as natural molecule.

According to another variant, the targeting sequences may be sequences encoding an N-terminal signal peptide ("prepeptide"), optionally in association with a signal responsible for retaining the protein in the endoplasmic reticulum (KDEL-type signal), or a vacuolar targeting signal or "propeptide". The presence of the N-terminal signal peptide or prepeptide allows the penetration of the nascent protein into the endoplasmic reticulum where a certain amount of post-translational processing occurs, particularly the cleaving of the signal peptide, the N-glycosylations, if the protein in question has N-glycosylation sites, and the formation of disulfide bridges. Among these various signals, the prepeptide responsible for the targeting of the protein into the endoplasmic reticulum, is dominant. It is normally a hydrophobic N-terminal signal peptide having between 10 and 40 amino acids and being of animal or plant origin. Preferably, it is a prepeptide of plant origin, for example that of sporamine, barley lectin, plant extensin,  $\alpha$ -mating factor, pathogenesis protein 1 or 2.

Normally, the signal peptide is cleaved by a peptidase signal upon the co-translational introduction of the nascent polypeptide into the lumen of the RER. The mature protein no longer contains this N-terminal extension.

The targeting sequences can, besides the prepeptide, also comprise an endoplasmic retention signal, consisting of the KDEL, SEKDEL or HEKDEL peptides. These signals normally exist at the C-terminal end of the protein and remain on the mature protein. The presence of this signal tends to increase the recombinant protein yields.

The targeting signals may, besides the prepeptide, also comprise a vacuolar targeting signal or "propeptide". In the presence of such a signal, after passing into the RER, the protein is targeted toward the vacuoles of the aqueous tissues, the leaves for example, as well as to the protein bodies of the storage tissues, for example the seeds, tubers and roots. The targeting of the protein toward the protein bodies of the seed is particularly advantageous because of the capacity of the seed to accumulate proteins, up to 40% of the proteins relative to the dry matter, in cellular organelles derived from the vacuoles, called protein bodies and because of the possibility of stocking, for several years, the seeds containing the recombinant proteins in the dehydrated state.

As propeptide, it is possible to use a signal of animal or plant origin, the plant signals being particularly preferred, for example prosporamine. The propeptide may be N-terminal ("N-terminal targeting peptide" or NTTP), or C terminal (CTTP). Since the propeptides are normally cleaved upon entry of the protein into the vacuole, it is not present in the mature protein.

The use of the signal peptide or prepeptide can lead to the glycosylation of the protein. Normally, globin has no N-glycosylation sites, but these may be introduced by mutagenesis. The  $\alpha$  and  $\beta$  chains can also have O-glycosylation sites.

In the absence of any targeting signal, the protein is expressed in the cytoplasm.

The nucleic acid introduced into the plant cell may also comprise sequences for regulation of transcription which are recognized by the plant cell. The nucleic acid is in this case a "chimeric gene". The regulatory sequences comprise one or more promoter(s) of plant or viral origin or obtained from Agrobacterium tumefaciens. They may be constitutive promoters, for example the CaMV 35S, the double 35S, the Nos or OCS promoters, or promoters specific for certain tissues such as the grain or specific for certain phases of development of the plant. As promoters specific for seeds, there may be mentioned the promoter of the gene for napin and for the acyl carrier protein (ACP) (EP-A-0,255,378), as well as the promoters of the AT2S genes of Arabidopsis thaliana, that is to say the PAT2S1, PAT2S2, PAT2S3 and PAT2S4 promoters (Krebbers et al., Plant Physiol., 1988, vol. 87, pages 859-866). It is

particularly preferable to use the cruciferin or phaseolin promoter or pGEA1 and pGEA6 of Arabidopsis, promoters of genes of the "em, Early Methionine labeled protein" type, which is strongly expressed during the phases of drying of the seed.

It is possible to envisage using "enhancers" to improve the efficiency of expression. When the transformation occurs directly in the chloroplast and mitochondrial genomes, gene promoters specific for these compartments can be used.

The sequences for regulation of transcription normally comprise sequences for termination of transcription which are of plant or of viral origin, for example 35S, or of bacterial origin (Agrobacterium).

When the transforming nucleic acid does not comprise regulatory sequences, it is preferable to add onto each end of the nucleic acid a DNA sequence homologous to the genomic sequences which are adjacent to a specific insertion site in the genome. This allows the integration of the construct by homologous recombination, at a site where endogenous regulatory sequences can control the expression of the heterologous sequences.

The nucleic acids of the invention may also comprise one or more intron(s), preferably of plant origin. These introns, which are obtained from a plant gene, are introduced artificially so as to increase the efficiency of expression of the heterologous sequence.

Indeed, it has been demonstrated, particularly in monocotyledonous plants, that the insertion of an intron into the untranslated 5' portion of a gene, that is to say between the site of initiation of transcription and the site of initiation of translation, leads to an improvement in the stability of the messenger, and consequently, to a better expression. The intron(s) used in this manner are obtained preferably from a monocotyledonous plant such as maize. This is preferably, but not necessarily, the first intron of the gene.

The nucleic acid sequence encoding  $\alpha$ - and  $\beta$ -globin and its variants is normally cDNA. Appropriate sequences are illustrated in Figures 2 and 3, any degenerate sequence can also be used as well as the sequences of the variants as defined above.

The introduction of a nucleic acid molecule(s) into the plant cell can be carried out in a stable manner either by transformation of the nuclear genome, or by transformation of the chloroplast genome of the plant cell, or by transformation of the mitochondrial genome.

For the transformation of the nuclear genome, conventional techniques may be used. All known means for introducing foreign DNA into plant cells may be used, for example Agrobacterium, electroporation, protoplast fusion, particle gun bombardment, or penetration of DNA into cells such as pollen, microspore, seed and immature embryo. Viral vectors such as the Gemini viruses or the satellite viruses may also be used as introducing means. Agrobacterium tumefaciens and rhizogenes constitute the preferred means. In this case, the sequence of the invention is introduced into an appropriate vector with all the necessary regulatory sequences such as promoters, terminators and the like, as well as any sequence necessary for selecting the transformants which have integrated the heterologous sequences.

The transformation of the nuclear genome of the plant cell is often carried out using the targeting signals mentioned above and which determine the cellular compartment where the expression and/or accumulation of the protein will occur.

According to another variant of the invention, the introduction of the nucleic acid into the plant cell can be carried out by the transformation of the mitochondrial or chloroplast genomes (see for example Carrer et al., Mol. Gen. Genet., 1993, 241, 49-56).

Techniques for direct transformation of the chloroplasts or the mitochondria are known per se and may comprise the following steps:

- i) introducing transformant DNA by the biolistic technique (Svab et al., P.N.A.S., 1990, 87, 8526-8530);
- ii) integrating the transformant DNA by two homologous recombination events;
- iii) selectively removing copies of the wild-type genome during repeated cell divisions on selective medium.

In order to allow the homologous recombination of the transformant DNA, two DNA fragments homologous to the genomic sequences, for example the

rbcL and ORF 512 genes are added to each end of the DNA to be inserted into the genome.

The direct transformation of the chloroplasts or mitochondria has the advantage of substantially increasing the yield of hemoglobin but the N-terminal methionine is retained.

According to another variant of the invention, the heterologous nucleic acid can be introduced into the plant cell by means of a viral vector.

The method of the invention comprises a step of detecting the hemin proteins and in particular hemoglobin and its derivatives. This makes it possible to verify if the plant or the plant cell is capable, not only of expressing the heterologous proteins, but also of assembling them correctly with the porphyrin nucleus. For the hemoglobin in a complex environment containing other chromophores or molecules which scatter light, detection by time-resolved optical spectroscopy will be advantageously used. This technique is described in detail in the examples. Other detection techniques consist in using antibodies specific for the alpha or beta globin chains or their variants. The spectrometric and immunological techniques can be used in association with each other. The use of these techniques makes it possible to select the plants which are capable of producing hemoglobin and its derivatives according to the invention.

The method of the invention comprises, in addition, a step of recovering or extracting hemoglobin or its derivatives from plant tissues. The extraction is normally carried out by grinding the tissues, for example leaves or grains, in an appropriate buffer, filtering the ground product, precipitating the proteins in the supernatant, centrifuging and taking up the pellet in an appropriate buffer with dialysis. A partial purification step can also be carried out at this stage by chromatography on an anion-exchange column.

The tetramer of hemoglobin, or of its derivatives, is purified by two successive chromatographies on an ion-exchange resin followed by a step of concentrating and saturating the concentrate with carbon monoxide. These techniques are described in detail in the examples.

When the expression of hemoglobin and of its derivatives take place under the control of a constitutive promoter, such as the 35S double promoter, an expression level of at least 1% hemoglobin compared with the total proteins may be obtained. The proteins represent about 10% of the dry mass of the leaf and a ton of dry tobacco leaves is harvested per hectare. It is therefore possible to obtain of the order of 100 grams of hemoglobin per hectare of tobacco cultivated, assuming that only 10% of the hemoglobin produced is purified.

The method of the invention therefore allows the production of hemoglobin at very low costs with a higher production capacity than that obtained using fermenters of the culture of bacteria or yeast.

Besides the method of transformation, the invention also includes vectors comprising one or more nucleic acid(s) or chimeric gene(s) defined above. As an example of vectors, there may be mentioned binary vectors or plasmids, viral vectors such as gemini viruses or the CaMVs.

The invention also relates to the plant cells transformed with the nucleic acid sequences of the invention. Preferably, they are transformed plant cells capable of producing one or more hemoglobin(s) or derivatives of hemoglobin according to the invention.

They may be plant cell cultures in vitro, for example in liquid medium. Various modes of culture ("batch", "fed batch" or continuous) for this type of cells are currently under study. The "batch" cultures are comparable to those carried out in an Erlenmeyer flask since the medium is not changed, these cells thus have only a limited quantity of nutrient materials. The "fed batch" culture corresponds, for its part, to a "batch" culture with programmed supply of substrate. For a continuous culture, the cells are supplied continuously with nutrient medium. An equal volume of the biomass-medium mixture is removed in order to maintain the volume of the reactor constant. The quantities of plant biomass which can be envisaged with cultures in bioreactors are variable depending on the plant species, the mode of culture and the type of bioreactor. Under certain conditions, biomass densities of about 10 to 30 g of dry weight per liter of culture can be obtained for species such as Nicotiana tabacum, Vinca rosea and Catharanthus roseus.

The cells of the invention can also be immobilized, which makes it possible to obtain a constant and prolonged production of hemoglobin. The separation of the hemoglobin and the plant biomass is also facilitated. As immobilization method, there may be mentioned immobilization in alginate or agar beads, inside polyurethane foam, or alternatively inside hollow fibers.

The cells of the invention may also be root cultures. The roots cultivated in vitro, in a liquid medium, are called "Hairy roots", they are roots transformed by the bacterium Agrobacterium rhizogenes.

Instead of producing the hemoglobin of the invention by culturing plant cells, it is possible to regenerate chimeric or transgenic plants from transformed explants, using techniques known per se.

As appropriate plants, there may be mentioned the Angiospermae comprising monocotyledonous and dicotyledonous plants. More particularly, there may be mentioned tobacco, species belonging to botanic families such as leguminous plants (for example beans, peas and the like), cruciferous plants (for example cabbage, radish, rapeseed and the like), Solanaceae (for example tomatoes, potato and the like), Cucurbitaceae (for example melon), Chenopodiaceae (for example beetroot), Umbelliferae (for example carrots, celery and the like). There may also be mentioned cereals such as wheat, maize, barley, triticale and rice, oleaginous plants such as sunflower and soybean. Tobacco, potato, tomato and maize are particularly preferred. For potato, the expression takes place preferably in the tubers.

The invention also relates to the seeds of transgenic plants capable of producing hemoglobin as well as their progeny.

The invention also relates to the hemin proteins which may be obtained by the method of the invention, in particular, the hemin proteins capable of reversibly binding oxygen, for example the hemoglobins and derivatives thereof.

The hemoglobins of the invention are capable of binding O<sub>2</sub> in a reversible manner with an affinity (P<sub>50</sub>) preferably close to physiological values (37°C, pH 7.40). The affinity of the molecule for O<sub>2</sub> is expressed as P<sub>50</sub>: that is to say the partial pressure of O<sub>2</sub> when hemoglobin or its derivatives is 50% saturated. The P<sub>50</sub> is measured according to the usual techniques, for example by means of an analyzer which

measures the percentage O<sub>2</sub> saturation as a function of the O<sub>2</sub> pressure (Kister et al., 1987). Normally, the hemoglobins of the invention have an acceptable autoxidation rate in order to minimize the formation of methemoglobin which is unsuited to the transport of O<sub>2</sub>. This characteristic can be measured by the absorption spectrum.

Preferably, the hemoglobins of the invention are tetramers, preferably alpha<sub>2</sub> beta<sub>2</sub>, beta<sub>4</sub>, or optionally tetramers of chimeric α/β subunits (Dumoulin et al., 1994, Art. Cells, Blood Subst., and Immob. Biotech., 22, 733-738) or multiples of four subunits. The physical size of the complex should be at least that of the tetramer in order to avoid its filtration by the kidneys.

The hemin proteins of the invention can be used in numerous pharmaceutical, cosmetic or industrial applications. The invention relates in particular to pharmaceutical compositions comprising one or more hemin protein(s) according to any one of Claims 15 to 23, in association with a physiologically acceptable excipient.

In the pharmaceutical field, all the conditions requiring an improvement of the transport of oxygen can be treated with the hemoglobins of the invention, these conditions comprising the following:

- acute or chronic hemorrhage,
- states of shock,
- coronary or sylvian angioplasties,
- treatments of solid tumors, sensitization to gamma-therapy,
- preservation of organs before transplant and during transport,
- malignant hemopathies.

The hemoglobins of the invention are normally used in the form of an injection in solutions optionally stabilized as regards the tetrameric form of the complex (for example addition of pyridoxal phosphate or diaspirin) as regards autoxidation. It is also possible to use suspensions of hemoglobin grafted on a support in order to increase the lifetime in the bloodstream. The support may be any conventional support in this domain, for example polysaccharides.

Various aspects of the invention are illustrated in the figures:

- Figure 1: Iron-containing protoporphyrin III (IX),

- Figure 2: cDNA sequence of human  $\alpha$ -globin (423 base pairs), and corresponding protein,
  - Figure 3: cDNA sequence of human  $\beta$ -globin (438 base pairs), and corresponding protein,
  - Figure 4: Experimental device for flash photolysis. A pulsed laser serves for the photodissociation of the Hb ligands:  $(HbCO \rightarrow Hb + CO)$ . A second optical beam, oriented at  $90^\circ$ , detects changes in absorption as a function of time after dissociation.
  - Figure 5: Kinetics of bimolecular recombination of CO with hemoglobin in plant extract. The two phases correspond to the two allosteric states of Hb: R (rapid) and T (slow).
- Conditions: 0.1 atm CO, pH 6-6, 25°C, about 50% dissociation, 0-1, 1 and 10  $\mu M$  Hb.
- Figure 6: Kinetics of recombination of CO with hemoglobin as a function of the percentage dissociation (by variation of the laser energy). The kinetics are sensitive to the number of ligands dissociated (1 to 4). At a high level of dissociation, Hb (deoxy or mono-ligand containing) shifts toward the slow form "T". At a low laser energy, the tetramers (mainly with three ligands) remain in the rapid form "R".
  - Figure 7: Western-blot analysis of the extract of the seeds of the transgenic tobacco T26-22 transformed with the plasmid pBIOC59. The extracts of seeds (75  $\mu g$  of proteins) of a nontransformed tobacco (1) and of a transgenic tobacco (T26-22) (2), molecular weight markers (3) and HbA (50 ng) (4) are separated by SDS-PAGE 17% electrophoresis under reducing conditions. The Western blotting is carried out under the conditions described in section X.a. The molecular weight markers and the  $\alpha$  and  $\beta$  globins are indicated.
  - Figure 8: Western-blot analysis of the fractions obtained during partial purification. The proteins in the fractions eluted from Sephadryl S-100 (37  $\mu g$ ), from S-Sepharose (30  $\mu g$ ) which are obtained during purification from mixtures of control seeds [(2) and FE-Control (4) respectively] and of seeds accumulating rHb [(3) and FE-rHb (5) respectively], HbA (50 ng) (1) and molecular weight markers (6) were separated by SDS-PAGE 17% electrophoresis under reducing conditions. The Western

blotting is carried out under the conditions described in section X.a. The molecular weight markers and the  $\alpha$  and  $\beta$  globins are indicated.

- Figure 9: Kinetics of recombination of CO with the FE-rHb fraction. The kinetics, following flash photolysis, is characteristic of the normal tetrameric Hb. The FE-Control fraction obtained from the control plants gives a signal of amplitude 1 mOD, that is to say about 50 times weaker than that observed for the FE-rHb fraction (48 mOD).

- Figure 10: Kinetics of recombination of CO with the FE-rHb fraction at various laser intensity levels. Similar results are observed for HbA (Figure 6).

- Figure 11: Demonstration of the reversible binding of oxygen to the FE-rHb sample. Since the oxyhemoglobin samples give only weak signals, we used, for these measurements, the techniques of mixing a CO atmosphere and O<sub>2</sub>. After photodissociation of CO, the rapid phase corresponds to the binding of oxygen. The oxygen is then replaced with CO which can again be photodissociated. The figure also shows the kinetics of recombination of the CO of the same sample equilibrated under 1 atm or 0.1 atm CO.

## EXAMPLES

### I. CONSTRUCTION OF BASAL EXPRESSION BINARY PLASMIDS ALLOWING THE PRODUCTION OF RECOMBINANT PROTEINS IN TOBACCO LEAVES.

The expression of genes in tobacco leaves requires the following regulatory sequences:

- 1) the constitutive double 35S promoter (pd35S) of CaMV (cauliflower mosaic virus). It corresponds to a duplication of the transcription-activating sequence situated upstream of the TATA element of the natural 35S promoter (Kay et al., 1987).
- 2) the sequence for termination of transcription, 35S polyA terminator, which corresponds to the noncoding 3' region of the sequence of the circular

double-stranded DNA cauliflower mosaic virus producing the 35S transcript (Franck et al., 1980).

The constructions of the various plasmids via the use of recombinant DNA techniques (Sambrook et al., 1989) are derived from pBIOC4. This binary plasmid is derived from pGA492 (An, 1986) which contains, between the right and left borders derived from the plasmid pTiT37 of Agrobacterium tumefaciens, on its transfer DNA, the following sequences:

- the constitutive promoter of the nos gene encoding nopaline synthase (Depicker et al., 1982),
- the coding sequence of the nptII gene encoding neomycin phosphotransferase II (Berg and Berg, 1983) deleted off the region of the first 8 codons including the ATG methionine initiation codon and fused to the sequence of the first 14 codons of the coding sequence of the nos gene (Depicker et al., 1982), the coding sequence of the nos gene lacking the region of the first 14 codons, the nos terminator (Depicker et al., 1982), a polylinker (HindIII-XbaI-SacI-HpaI-KpnI-ClaI-BglII) preceding the cat gene encoding chloramphenicol acetyltransferase (Close and Rodriguez, 1982) and the termination sequences of gene 6 of the plasmid pTiA6 Agrobacterium tumefaciens (Liu et al., 1993).

To remove virtually the whole of the coding sequence of the cat gene, the plasmid pGA492 was doubly digested with SacI (restriction site of the polylinker) and with ScaI (restriction site present in the sequence of the cat gene) and then subjected to the action of the enzyme T4 DNA polymerase (New England Biolabs) according to the manufacturer's recommendations. The ligation of the modified plasmid (20 ng) was carried out in a reaction medium of 10 µl containing 1 µl of 10x T4 DNA ligase buffer (Amersham) and 2.5 U of T4 DNA ligase enzyme (Amersham) at 14°C for 16 hours. The E. coli DH5 $\alpha$  bacteria previously made competent were transformed (Hanahan, 1983).

The plasmid DNA of the clones obtained, selected on 12 µg/ml tetracycline, was extracted according to the alkaline lysis method (Birnboim and Doly, 1979) and analyzed by enzymatic digestion with restriction enzymes. Next, the HindIII restriction site of the plasmid DNA of the selected clone was modified at an EcoRI

restriction site with the aid of a phosphorylated HindIII - EcoRI adaptor (Stratagene Cloning Systems). To carry out this modification, 500 ng of plasmid DNA of the selected clone were digested with HindIII, dephosphorylated with the enzyme calf intestinal alkaline phosphatase (Boeringer Mannheim) according to the manufacturer's recommendations and coprecipitated in the presence of 1500 ng of HindIII-EcoRI adaptor DNA, 1/10 volume of 3 M sodium acetate pH 4.8 and 2.5 volumes of absolute ethanol at -80°C for 30 min. After centrifugation at 12000 g for 30 min, the precipitated DNA was washed with 70% ethanol, dried, taken up in 8 µl of water, heated at 65°C for 10 min, and then ligated in the presence of 1 µl of 10x T4 DNA ligase buffer (Amersham) and 2.5 U of the enzyme T4 DNA ligase (Amersham) at 14°C for 16 hours. After inactivation of the T4 DNA ligase at 65°C for 10 min, the ligation reaction mixture was digested with EcoRI, purified by electrophoresis on a 0.8% agarose gel, electroeluted (Sambrook et al., 1989), precipitated in the presence of 1/10 volume of 3 M sodium acetate pH 4.8 and 2.5 volumes of absolute ethanol at -80°C for 30 min, centrifuged at 12000 g for 30 min, washed with 70% ethanol and then dried. The E. coli DH5 $\alpha$  bacteria previously made competent were transformed (Hanahan, 1983). The plasmid DNA of the clones obtained, selected on 12 µg/ml tetracyclin, was extracted according to the alkaline lysis method (Birnboim and Doly, 1979) and analyzed by enzymatic digestion with HindIII and EcoRI in particular. The resulting binary plasmid, which now possesses only the last 9 codons of the coding sequences of the cat gene and in which the EcoRI site is unique, was called pBIOC4.

a. CONSTRUCTION OF THE EXPRESSION BINARY PLASMID pBIOC21.

The expression cassette, consisting of the pd35S promoter and the 35S polyA terminator, was isolated from the plasmid pJIT163D. The plasmid pJIT163D is derived from the plasmid pJIT163 which is itself derived from the plasmid pJIT60 (Guerineau and Mullineaux, 1993). The plasmid pJIT163 possesses an ATG codon between the HindIII and Sall sites of the polylinker. To eliminate this ATG and to obtain the plasmid pJIT163D, the plasmid DNA pJIT163 was doubly digested with HindIII and Sall, purified by electrophoresis on a 0.8% agarose gel, electroeluted

(Sambrook et al., 1989), precipitated in the presence of 1/10 volume of 3 M sodium acetate pH 4.8 and 2.5 volumes of absolute ethanol at -80°C for 30 min, centrifuged at 12000 g for 30 min, washed with 70% ethanol, dried, subjected to the action of the Klenow enzyme (New England Biolabs) according to the manufacturer's recommendations, deproteinized by extraction with 1 volume of phenol:chloroform:isoamyl alcohol (25:24:1) and then 1 volume of chloroform:isoamyl alcohol (24:1), precipitated in the presence of 1/10 volume of 3 M sodium acetate pH 4.8 and 2.5 volumes of absolute ethanol at -80°C for 30 min, centrifuged at 12000 g for 30 min, washed with 70% ethanol, dried and finally ligated in the presence of 1 µl of 10x T4 DNA ligase buffer (Amersham) and 2.5 U of T4 DNA ligase enzyme (Amersham) at 14°C for 16 hours. The E. coli DH5 $\alpha$  bacteria previously made competent, were transformed (Hanahan, 1983). The plasmid DNA of the clones obtained, selected on 50 µg/ml ampicillin, was extracted according to the alkaline lysis method (Birnboim and Doly, 1979) and analyzed by enzymatic digestion with restriction enzymes. To isolate the expression cassette consisting of the pd35S promoter and of the 35S polyA terminator (SacI-XhoI fragment), the plasmid DNA of the pJIT163D clone selected was digested with SacI and XhoI. The SacI-XhoI- fragment, carrying the expression cassette, was purified by electrophoresis on a 0.8% agarose gel, electroeluted (Sambrook et al., 1989) precipitated in the presence of 1/10 volume of 3 M sodium acetate pH 4.8 and 2.5 volumes of absolute ethanol at -80°C for 30 min, centrifuged at 12000 g for 30 min, washed with 70% ethanol, dried and then subjected to the action of Mung Bean nuclease enzyme (New England Biolabs) according to the manufacturer's recommendations. This purified insert (200 ng) was cloned into the plasmid DNA of pBIOC4 (20 ng) digested with EcoRI, treated with the Mung Bean nuclease enzyme and dephosphorylated with the enzyme calf intestinal alkaline phosphatase (Boehringer Mannheim) according to the manufacturer's recommendations. The ligation reaction was carried out in 20 µl, in the presence of 2 µl of 10x T4 DNA ligase buffer (Amersham), 2 µl of 50% polyethylene glycol 8000 and 5 U of T4 DNA ligase enzyme (Amersham) at 14°C for 16 hours. The E. coli DH5 $\alpha$  bacteria previously made competent were transformed (Hanahan, 1983). The plasmid DNA of the clones obtained, selected on 12 µg/ml tetracyclin was extracted according

to the alkaline lysis method (Birnboim and Doly, 1979) and analyzed by enzymatic digestion with restriction enzymes. The resulting plasmid was called pBIOC21.

b. CONSTRUCTION OF THE CO-EXPRESSION BINARY PLASMID pBIOC43.

The co-expression binary plasmid will allow expression of two genes in the same binary vector.

The co-expression binary plasmid is derived from pBIOC21. It contains two expression cassettes each consisting of a pd35S promoter and a 35S polyA terminator but differ in the polylinker separating the promoter from the terminator. One of the expression cassettes is that of pBIOC21 already described in paragraph I.a. The other expression cassette was obtained by replacing the HindIII-BamHI-SmaI-EcoRI polylinker of pJIT163D (described in paragraph I.a.) by a HindIII-EcoRI adaptor carrying the PacI, Ascl, MluI and HpaI restriction sites. This adaptor was obtained by renaturation of the 2 oligodeoxynucleotides WD11 (5' AGC TGA TTA ATT AAG GCG CGC CAC GCG TTA AC 3') and WD12 (5' AAT TGT TAA CGC GTG GCG CGC CTT AAT TAA TC 3') which are complementary for their 28 terminal 3' nucleotides. One hundred  $\mu$ M of each of these two oligodeoxynucleotides were previously phosphorylated by the action of 10 U of T4 polynucleotide kinase enzyme (New England Biolabs) in a total reaction volume of 10  $\mu$ l containing 1  $\mu$ l of 10x T4 polynucleotide kinase buffer (New England Biolabs) and 3  $\mu$ l of ATP (95 mM). The two reaction mixtures were incubated at 37°C for 1 hour, and then at 65°C for 20 min. They were then combined and their volume was adjusted to 500  $\mu$ l. After extraction with 1 volume of phenol:chloroform:isoamyl alcohol (25:24:1) and 1 volume of chloroform:isoamyl alcohol (24:1), 50  $\mu$ l of 3 M sodium acetate pH 6.0 were added. The reaction mixture was incubated at 80°C for 10 min and then cooled slowly to room temperature. The DNA was then precipitated in the presence of 2.5 volumes of absolute ethanol at -80°C for 30 min, centrifuged at 14000 g at 4°C for 1 hour, washed with 70% ethanol, centrifuged at 14000 g at 4°C for 10 min, dried, taken up in 10  $\mu$ l of H<sub>2</sub>O. The HindIII-EcoRI DNA fragment was then cloned at the HindIII-EcoRI sites of the

plasmid DNA pJIT163D previously dephosphorylated with the enzyme calf intestinal alkaline phosphatase (New England Biolabs) according to the manufacturer's recommendations. The ligation reaction was carried out in a reaction volume of 20  $\mu$ l in the presence of 1 U of T4 DNA ligase (Gibco-BRL) for a total DNA concentration of 8.5 nM with a vector/insert molar ratio of 1 and of 4  $\mu$ l of 5x T4 DNA ligase buffer (Gibco-BRL) at 25°C for 16 hours. The E. coli DH5 $\alpha$  bacteria previously made competent were transformed (Hanahan, 1985). The plasmid DNA of the clones obtained, selected on 100  $\mu$ g/ml ampicillin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC42. Its validity was verified by sequencing with the aid of the "Sequenase Version 2.0 DNA Sequencing" kit marketed by United States Biochemical (USB) according to the dideoxynucleotides method (Sanger et al., 1977). The reaction conditions follow the manufacturer's recommendations except for the denaturation and hybridization. The reaction medium containing the plasmid DNA (0.5 to 1 pmol), the oligonucleotide primer (2 pmol), 10% DMSO and the 1x reaction buffer (USB), is incubated at 100°C for 10 min, then suddenly cooled to -80°C in dry ice.

From pBIOC42, the DNA fragment encoding the expression cassette consisting of the pd35S promoter and of the 35S polyA terminator was isolated by double digestion with SacI and XhoI. It was purified by electrophoresis on a 0.75% agarose gel, and then subjected to the action of the "Geneclean II" kit marketed by BIO101 according to the manufacturer's recommendation. Next, this DNA fragment was inserted at the SacI and XhoI sites of the plasmid pBCSK+ marketed by Stratagene and previously dephosphorylated with the enzyme calf intestinal alkaline phosphatase (New England Biolabs) according to the manufacturer's recommendations. The ligation was carried out in a reaction volume of 20  $\mu$ l in the presence of 1 U of T4 DNA ligase (Gibco-BRL) for a total DNA concentration of 8.5 nM with a vector/insert molar ratio of 1 and of 4  $\mu$ l of 5x T4 DNA ligase buffer (Gibco-BRL) at 25°C for 16 hours. The E. coli DH5 $\alpha$  bacteria previously made competent were transformed (Hanahan, 1985). The plasmid DNA of the clones obtained, selected on 30  $\mu$ g/ml of chloramphenicol, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed

by enzymatic digestion with restriction enzymes. The resulting plasmid was called pBIOC75.

From pBIOC75, the DNA fragment carrying the expression cassette consisting of the pd35S promoter and the 35S polyA terminator was isolated by digestion with KpnI. It was purified by electrophoresis on a 0.75% agarose gel, and then subjected to the action of the "Geneclean II" kit marketed by BIO101 according to the manufacturer's recommendations. Next, this DNA fragment was ligated to the plasmid DNA of pBIOC21 digested with KpnI and dephosphorylated with the enzyme calf intestinal alkaline phosphatase (New England Biolabs) according to the manufacturer's recommendations. The ligation was carried out in a reaction volume of 20 µl in the presence of 1 U of T4 DNA ligase (Gibco-BRL) for a total DNA concentration of 8.5 nM with a vector/insert molar ratio of 1 and of 4 µl of 5x T4 DNA ligase buffer (Gibco-BRL) at 25°C for 16 hours. The E. coli DH5 $\alpha$  bacteria previously made competent were transformed (Hanahan, 1985). The plasmid DNA of the clones obtained, selected on 12 µg/ml of tetracycline, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion with restriction enzymes. The resulting plasmid was called pBIOC43.

## **II. CONSTRUCTION OF THE CHIMERIC GENES ENCODING THE $\alpha$ AND $\beta$ GLOBIN CHAINS ALLOWING EXPRESSION OF RECOMBINANT HUMAN HEMOGLOBIN IN THE CYTOPLASM OF TOBACCO LEAVES.**

The plasmid alpha1pJW101 contains the cDNA for the  $\alpha$  globin chain, cloned into the plasmid pMB9 as described by Wilson et al. (1978).

The M13mp10 phage cIIIX beta contains the cDNA for the  $\beta$  globin chain, cloned into the M13 mp10 phage as described by Nagai et al. (1985). In this construction, the cDNA encoding  $\beta$  globin was inserted in 3' of the coding sequence for the cII protein of the lambda phage, followed by that encoding the FX tetrapeptide, forming a fusion gene in which the initiator ATG codon of  $\beta$  globin has been deleted.

a. CONSTRUCTION OF THE PLASMID pBIOC44 CONTAINING THE cDNA ENCODING  $\alpha$  GLOBIN FOR CYTOPLASMIC TARGETING.

To obtain cytoplasmic targeting of the  $\alpha$  globin chain, the initiator methionine codon of the  $\alpha$  globin chain was maintained.

The cDNA encoding the cytoplasmic targeting  $\alpha$  globin chain was obtained in three stages. The first two stages made it possible to suppress the internal HindIII site (substitution of a T for a C) at position 276 of the coding sequence whereas the third stage combines the 2 cDNA fragments encoding the recombinant  $\alpha$  globin chain.

The first stage consisted in the amplification of the first 95 codons of the mature  $\alpha$  globin chain on the plasmid alpha1pJW101 with the aid of the 2 oligodeoxynucleotides, WD13 (5' tacaagcttaaca ATG GTG CTG TCT CCg GCC GAC 3') and AD27 (5' CGG GTC CAC CCG GAG CTT GTG 3'). The WD13 primer provides the HindIII restriction site, the sequence aaca favoring the initiation of translation (Joshi, 1987) and preceding the initiator ATG codon followed by the first 6 codons of the mature  $\alpha$  globin chain of which the fourth (CCT) is substituted for CCg (silent mutation) in order to create the EagI restriction site. The AD27 primer allows the suppression of the HindIII restriction site by substitution of nucleotide T for C (position 276 of the coding sequence). The PCR amplification was carried out in 100  $\mu$ l of reaction medium containing 10  $\mu$ l of 10x Taq DNA polymerase buffer (100 mM Tris-HCl pH 8.4, 500 mM KCl and 20 mM MgCl<sub>2</sub>), 16  $\mu$ l of the dNTP mixture (1.25 mM dATP, 1.25 mM dCTP, 1.25 mM dGTP and 1.25 mM dTTP), 10  $\mu$ l of each of the primers described above at 10  $\mu$ l, 10  $\mu$ l of template DNA (alpha1pJW101) at 1 ng/ $\mu$ l and 0.5  $\mu$ l of Taq DNA polymerase at 5 U/ $\mu$ l (Perkin Elmer). Thirty cycles each comprising 30 sec of denaturation at 97°C, 1 min hybridization at 55°C and 2 min extension at 72°C were carried out in the Appligène "Crocodile II" apparatus. The amplified DNA fragments were then purified by electrophoresis on a 1.8% agarose gel and by the action of the "Geneclean II" kit marketed by BIO101 according to the

manufacturer's recommendations. The purified amplified DNA fragments are taken up in 20  $\mu$ l.

The second stage consisted in the amplification of the last 54 codons of the mature  $\alpha$  globin chain on the plasmid alpha1pJW101 with the aid of the 2 oligodeoxynucleotides, AD26 (5' CAC AAG CTC CGG GTG GAC CCG 3') and WD14 (5' gcgaattc TCA ACG GTA TTT GGA GGT CAG CAC 3'). The WD14 primer provides the EcoRI restriction site situated just after the stop codon. The AD26 primer allows the suppression of the HindIII restriction site by substitution of nucleotide T for C (position 276 of the coding sequence). The PCR amplification was carried out as described in the first stage. The treatment of the amplified DNA fragments was carried out as described in the first stage.

The third stage was the PCR amplification of the complete cDNA encoding the  $\alpha$  globin chain (142 codons including the initiator ATG). The two types of DNA fragments amplified in the first and second stages served as template DNA and the two primers used were WD13 and WD14. The PCR amplification was carried out as described in the first stage except that the hybridization temperature of the cycle is 60°C. The amplified DNA fragments were then extracted with H<sub>2</sub>O-saturated ether after having adjusted the volume to 500  $\mu$ l with TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). After extraction with 1 volume of phenol:chloroform:isoamyl alcohol (25:24:1) and 1 volume of chloroform:isoamyl alcohol (24:1), the DNA fragments were precipitated in the presence of 1/10 volume of 3 M sodium acetate pH 6.0 and 2 volumes of absolute ethanol at -80°C for 30 min, centrifuged at 14000 g at 4°C for 30 min, washed with 70% ethanol, centrifuged at 14000 g at 4°C for 10 min, dried, taken up in 50  $\mu$ l of H<sub>2</sub>O. Next, 25  $\mu$ l of these DNA fragments were doubly digested with HindIII and EcoRI, purified by electrophoresis on 1.8% agarose gel and by the action of the "Geneclean II" kit (BIO101) and cloned at the HindIII and EcoRI sites of the plasmid pNEB193 marketed by New England Biolabs, and previously dephosphorylated with the enzyme calf intestinal alkaline phosphatase (New England Biolabs) according to the manufacturer's recommendations. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 100  $\mu$ g/ml ampicillin, was extracted according to the

alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC44. The nucleotide sequence of the cDNA encoding the recombinant  $\alpha$  globin chain was verified by sequencing with the aid of the "Sequenase Version 2.0 DNA Sequencing" kit marketed by United States Biochemical (USB) as described in section I.b. The sequencing revealed two silent mutations situated at the forty-eighth nucleotide (C modified to T) and at the fifty-fourth (T modified to C) of the coding sequence for the  $\alpha$  globin chain.

b. CONSTRUCTION OF THE PLASMID pBIOC45 CONTAINING THE cDNA ENCODING  $\beta$  GLOBIN FOR CYTOPLASMIC TARGETING.

To obtain cytoplasmic targeting of the  $\beta$  globin chain, the methionine codon was fused with the first codon of the mature  $\beta$  globin chain by maintaining the open reading frame since ATG had been deleted from the construct M13mp10 cIIFX beta.

The cDNA encoding the cytoplasmic targeting  $\beta$  globin chain was obtained by PCR amplification of the 146 codons constituting the mature  $\beta$  globin chain on the phage M13mp10 cIIFX beta with the aid of the 2 oligodeoxynucleotides WD15 (5' gtcattaaatcca ATG GTG CAC CTG ACT CCT GAG GAG AAG TCg GCC GTT AC 3') and WD16 (5' aatgagctcgtaacgcgt TTA GTG ATA CTT GTG GGC CAG GGC 3'). The WD15 primer provides the PacI restriction site, the aaca sequence favoring the initiation of translation (Joshi, 1987) and the initiator ATG codon followed by the first 12 codons of the mature  $\beta$  globin chain of which the ninth (TCT) is substituted for TCg (silent mutation) in order to create the EagI restriction site. The WD16 primer provides the MluI, HpaI and SacI restriction sites placed after the stop codon. The PCR amplification and the treatment of the amplified DNA fragments were carried out as described in the third stage of section II.a. Next, 25  $\mu$ l of these DNA fragments were doubly digested with PacI and SacI, purified by electrophoresis on a 1.8% agarose gel and by the action of the "Geneclean II" kit (BIO101) and cloned at the PacI and SacI sites of the plasmid pNEB193 marketed by New England Biolabs, previously dephosphorylated by the enzyme calf intestinal alkaline phosphatase (New England

Biolabs) according to the manufacturer's recommendations. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 100 µg/ml ampicillin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC45. The nucleotide sequence of the cDNA encoding the recombinant β globin chain was verified by sequencing as described in section I.b.

c. CONSTRUCTION OF THE EXPRESSION BINARY PLASMIDS pBIOC46 AND pBIOC47, AND OF THE CO-EXPRESSION BINARY PLASMID pBIOC49 FOR CYTOPLASMIC TARGETING.

c.1. CONSTRUCTION OF THE BINARY PLASMID pBIOC46 CONTAINING cDNA ENCODING α GLOBIN FOR CYTOPLASMIC TARGETING.

Starting with pBIOC44, the HindIII-EcoRI fragment carrying the cDNA encoding the cytoplasmic targeting α globin chain was isolated by double enzymatic digestion with HindIII and EcoRI, purified by electrophoresis on a 1.8% agarose gel and by the action of the "Geneclean II" kit (BIO101). Next, this DNA fragment was ligated with the plasmid DNA of pBIOC21 doubly digested with HindIII and EcoRI, and dephosphorylated with the enzyme calf intestinal alkaline phosphatase (New England Biolabs) according to the manufacturer's recommendations. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 10 µg/ml tetracycline, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC46. The nucleotide sequence of the cDNA encoding the recombinant α globin chain was verified by sequencing as described in section I.b. The plasmid DNA of the binary vector pBIOC46 was introduced by direct transformation into the Agrobacterium tumefaciens LBA4404 strain according to the method of Holsters et al. (1978). The validity of the clone selected was verified by enzymatic digestion of the plasmid DNA introduced.

c.2. CONSTRUCTION OF THE BINARY PLASMID pBIOC47  
CONTAINING THE cDNA ENCODING  $\beta$  GLOBIN FOR CYTOPLASMIC  
TARGETING.

Starting with pBIOC45, the HindIII-EcoRI fragment carrying the cDNA encoding the cytoplasmic targeting  $\beta$  globin chain was isolated by double enzymatic digestion with HindIII (total digestion) and EcoRI (partial digestion), purified by electrophoresis on a 1.8% agarose gel and by the action of the "Geneclean II" kit (BIO101). Next, this DNA fragment was ligated with the plasmid DNA of pBIOC21 doubly digested with HindIII and EcoRI and dephosphorylated with the enzyme calf intestinal alkaline phosphatase (New England Biolabs) according to the manufacturer's recommendations. The ligation and the transformation were carried out as described in section I.b.

The plasmid DNA of the clones obtained, selected on 10  $\mu$ g/ml tetracycline, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBICO47. The nucleotide sequence of the cDNA encoding the recombinant  $\beta$  globin chain was verified by sequencing as described in section I.b. The plasmid DNA of the binary vector pBIOC47 was introduced by direct transformation into the Agrobacterium tumefaciens LBA4404 strain according to the method of Holsters et al. (1978). The validity of the clone selected was verified by enzymatic digestion of the plasmid DNA introduced.

c.3. CONSTRUCTION OF THE CO-EXPRESSING BINARY  
PLASMID pBIOC49 CONTAINING THE cDNA ENCODING THE  $\alpha$  AND  $\beta$   
GLOBINS FOR CYTOPLASMIC TARGETING.

The HindIII-EcoRI fragment carrying the cDNA encoding the cytoplasmic targeting  $\alpha$  globin chain was isolated from pBIOC44 described in section II.c.1., and ligated with the plasmid DNA of pBIOC43 doubly digested with HindIII and EcoRI, and previously dephosphorylated with the enzyme calf intestinal alkaline phosphatase (New England Biolabs) according to the manufacturer's

recommendations. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 12 µg/ml tetracycline, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC48.

The PacI-MluI fragment carrying the cDNA encoding the cytoplasmic targeting β globin chain was isolated from pBIOC45 described in section II.c.2., and ligated with the plasmid DNA of pBIOC48 doubly digested with PacI and MluI, and dephosphorylated by the enzyme calf intestinal alkaline phosphatase (New England Biolabs) according to the manufacturer's recommendations. The ligation and the transformation were carried out as described in section I.b., except that the E. coli Sure tet<sup>r</sup> strain was used in place of DH5α. The Sure tet<sup>r</sup> strain is derived from the Sure strain (Stratagene) made sensitive to tetracycline by the loss of the F' episome. The plasmid DNA of the clones obtained, selected on 10 µg/ml tetracycline, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC49.

The nucleotide sequence of the cDNAs encoding the recombinant α and β globin chains was verified by sequencing as described in section I.b. The plasmid DNA in the binary vector pBIOC49 was introduced by direct transformation into the Agrobacterium tumefaciens LBA4404 strain according to the method of Holsters et al. (1978). The validity of the clone selected was verified by enzymatic digestion of the plasmid DNA introduced.

### **III. CONSTRUCTION OF THE CHIMERIC GENES ENCODING THE α AND β GLOBIN CHAINS ALLOWING EXPRESSION OF RECOMBINANT HUMAN HEMOGLOBIN IN THE MITOCHONDRIA OF TOBACCO LEAVES**

To obtain mitochondrial targeting, the sequence encoding the transit peptide of the Nicotiana plumbaginifolia mitochondrial ATPase-F1 β subunit precursor (ATG GCT TCT CGG AGG CTT CTC GCC TCT CTC CTC CGT CAA TCG GCT CAA CGT GGC GGC GGT CTA ATT TCC CGA TCG TTA GGA AAC TCC ATC CCT AAA TCC GCT TCA CGC GCC TCT TCA CGC GCA TCC CCT AAG GGA

TTC CTC TTA AAC CGC GCC GTA CAG TAC) is fused with the first codon of the sequence encoding, on the one hand, the mature  $\alpha$  globin chain (deletion of the initiator ATG) and, on the other hand, the mature  $\beta$  globin chain (deletion of the initiator ATG) while maintaining the open reading frames.

The sequence encoding the Nicotiana plumbaginifolia mitochondrial ATPase F1  $\beta$  subunit is contained in the plasmid pTZ-catp2-1 provided by Boutry. This plasmid corresponds to the plasmid pTZ18R containing the cDNA (cNP10) as described by Boutry and Chua (1985).

The N-terminal transit peptide, composed of 54 amino acids as defined by Chaumont et al. (1994), was used during the carrying out of the constructions.

a. CONSTRUCTION OF THE PLASMID pBIOC50 CONTAINING THE cDNA ENCODING  $\alpha$  GLOBIN FOR MITOCHONDRIAL TARGETING.

To obtain mitochondrial targeting of the  $\alpha$  globin chain, the sequence encoding the transit peptide of the Nicotiana plumbaginifolia mitochondrial ATPase-F1  $\beta$  subunit precursor was fused with the first codon of the sequence encoding the mature  $\alpha$  globin chain while maintaining the open reading frame. The cleaving sequence between the sequences of the transit peptide and the mature  $\alpha$  globin chain is Tyr-Val.

The sequence encoding the transit peptide of the mitochondrial ATPase-F1  $\beta$  subunit precursor was amplified by PCR on the plasmid pTZ-capt2-1 with the aid of the 2 oligodeoxynucleotides, WD 17 (5' cgcaagcttaaca ATG GCT TCT CGG AGG CTT CTC 3') and WD18 (5' tag aat tC GGC cGG AGA CAG CAC GTA CTG TAC GGC GCG GTT TAA G 3'). The WD17 primer provides the HindIII restriction site, the aaca sequence promoting the initiation of translation (Joshi, 1987) and the first 7 codons of the transit peptide (including the initiator ATG). The WD18 primer provides the EcoRI restriction site, the first 5 codons of the sequence encoding the mature  $\alpha$  globin chain (an EagI restriction site is created by silent mutation in the fourth codon (CCT modified to CCg) and the last 7 codons of the sequence of the transit peptide. PCR amplification and the treatment of the amplified DNA fragments were carried out as described in the third step of chapter II.a. Next, these DNA fragments were doubly

digested with HindIII and EagI, purified by electrophoresis on a 1.8% agarose gel and by the action of the "Geneclean II" kit (BIO101) and cloned at the HindIII and EagI sites of the plasmid pBIOC44 described in section II.a., previously purified by electrophoresis on a 0.75% agarose gel and using the "Geneclean II" kit. The plasmid pBIOC44 was dephosphorylated by the enzyme calf intestinal alkaline phosphatase (New England Biolabs) according to the manufacturer's recommendations. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 100 µg/ml ampicillin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC50. The nucleotide sequence of this chimeric gene resulting from the translational fusion between the sequence encoding the transit peptide and the cDNA encoding the mature  $\alpha$  globin chain was verified by sequencing as described in section I.b. The sequencing revealed two silent mutations situated at the tenth nucleotide (C modified to A) and at the one hundred forty first (C modified to G) of the coding sequence for the transit peptide.

b. CONSTRUCTION OF THE PLASMID pBIOC51 CONTAINING THE cDNA ENCODING  $\beta$  GLOBIN FOR MITOCHONDRIAL TARGETING.

To obtain mitochondrial targeting of the  $\beta$  globin chain, the sequence encoding the transit peptide of the Nicotiana plumbaginifolia mitochondrial ATPase-F1  $\beta$  subunit precursor was fused with the first codon of the sequence encoding the mature  $\beta$  globin chain while maintaining the open reading frame. The cleaving sequence between the sequences of the transit peptide and the mature  $\beta$  globin chain is Tyr-Val.

The sequence encoding the transit peptide of the mitochondrial ATPase-F1  $\beta$  subunit precursor was amplified by PCR on the plasmid pTZ-catp2-1 with the aid of the 2 oligodeoxynucleotides, WD19 (5' gtcattaaatcca ATG GCT TCT CGG AGG CTT CTC GCC TCT C 3') and WD20 (5' aatgagct C GGC cGA CTT CTC CTC AGG AGT CAG GTG CAC GTA CTG TAC GGC GCG GTT TAA G 3'). The WD19 primer provides the PacI restriction site, the aaca sequence promoting the initiation of translation (Joshi, 1987) and preceding the first 9 codons of the transit peptide (including the initiator ATG). The WD20 primer provides the SacI restriction site, the

first 10 codons of the sequence encoding the mature  $\beta$  globin chain (an EagI restriction site is created by silent mutation in the ninth codon (TCT modified to TCg)) and the last 7 codons of the sequence of the transit peptide. The PCR amplification and the treatment of the amplified DNA fragments were carried out as described in the third stage of section II.a. Next, these DNA fragments were doubly digested with PacI and EagI, purified by electrophoresis on a 1.8% agarose gel and by the action of the "Geneclean II" kit (BIO101) and cloned at the PacI and EagI sites of the plasmid pBIOC45 described in section II.b., previously purified by electrophoresis on a 0.75% agarose gel and using the "Geneclean II" kit. The plasmid pBIOC45 was dephosphorylated by the enzyme calf intestinal alkaline phosphatase (New England Biolabs) according to the manufacturer's recommendations. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 100  $\mu$ g/ml ampicillin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC51. The nucleotide sequence of this chimeric gene resulting from the translational fusion between the sequence encoding the transit peptide and the cDNA encoding the mature  $\beta$  globin chain was verified by sequencing as described in section I.b.

c. CONSTRUCTION OF THE CO-EXPRESSION BINARY PLASMID pBIOC53  
CONTAINING THE cDNAs ENCODING THE  $\alpha$  AND  $\beta$  GLOBINS, FOR  
MITOCHONDRIAL TARGETING.

The HindIII-EcoRI fragment carrying the cDNA encoding the mitochondrial targeting  $\alpha$  globin chain was isolated from pBIOC50 described in section III.a., and ligated to the plasmid DNA of pBIOC43 doubly digested with HindIII and EcoRI, and dephosphorylated by the enzyme calf intestinal alkaline phosphatase (New England Biolabs) according to the manufacturer's recommendations. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 10  $\mu$ g/ml tetracycline, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC52.

The PacI-MluI fragment carrying the cDNA encoding the mitochondrial targeting  $\beta$  globin chain was isolated from pBIOC51 described in section III.b., and ligated to the plasmid DNA of pBIOC52 doubly digested with PacI and MluI, and dephosphorylated by the enzyme calf intestinal alkaline phosphatase (New England Biolabs). The ligation and the transformation were carried out as described in section II.c.3 using the *E. coli* Sure tet<sup>r</sup> strain. The plasmid DNA of the clones obtained, selected on 10  $\mu$ g/ml tetracycline, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC53.

The nucleotide sequence of the cDNAs encoding the recombinant  $\alpha$  and  $\beta$  globin chains allowing  $\alpha$  mitochondrial targeting was verified by sequencing as described in section I.b. The plasmid DNA of the binary vector pBIOC53 was introduced by direct transformation into the *Agrobacterium tumefaciens* LBA4404 strain according to the method of Holsters et al. (1978). The validity of the clone obtained was verified by enzymatic digestion of the plasmid DNA introduced.

#### **IV. CONSTRUCTION OF THE CHIMERIC GENES ENCODING THE $\alpha$ AND $\beta$ GLOBIN CHAINS ALLOWING EXPRESSION OF RECOMBINANT HUMAN HEMOGLOBIN IN THE CHLOROPLASTS OF TOBACCO LEAVES.**

To obtain chloroplastic targeting, the sequence encoding the transit peptide of the precursor of the small subunit of ribulose 1,5-diphosphate carboxylase of *Pisum sativum* L. (ATG GCT TCT ATG ATA TCC TCT TCA GCT GTG ACT ACA GTC AGC CGT GCT TCT ACG GTG CAA TCG GCC GCG GTG GCT CCA TTC GGC GGC CTC AAA TCC ATG ACT GGA TTC CCA GTT AAG AAG GTC AAC ACT GAC ATT ACT TCC ATT ACA AGC AAT GGT GGA AGA GTA AAG TGC) is fused with the first codon of the sequence encoding, on the one hand, the mature  $\alpha$  globin chain (deletion of the initiator ATG) and, on the other hand, the mature  $\beta$  globin chain (deletion of the initiator ATG) while maintaining the open reading frames.

This N-terminal transit peptide, composed of 57 amino acids, as defined by Anderson et al. (1986), was isolated from the plasmid pJIT117 (Guerineau et al., 1988) and used during the carrying out of the constructions.

a. CONSTRUCTION OF THE PLASMID pBIOC55 CONTAINING THE cDNA ENCODING  $\alpha$  GLOBIN FOR CHLOROPLAST TARGETING.

To obtain chloroplast targeting of the  $\alpha$  globin chain, the sequence encoding the transit peptide of the precursor of the small subunit of the ribulose 1,5-diphosphate carboxylase of *Pisum sativum* L. was fused with the first codon of the sequence encoding the mature  $\alpha$  globin chain while maintaining the open reading frame. The cleaving sequence between the sequences of the transit peptide and of the mature  $\alpha$  globin chain is Cys-Val.

The sequence of the transit peptide of the precursor of the small subunit of ribulose 1,5-diphosphate carboxylase was amplified by PCR on the plasmid pJIT117 with the aid of the 2 oligodeoxynucleotides, WD21 (5' cgcaagcttaaca ATG GCT TCT ATG ATA TCC TCT TCA GC 3') and WD22 (5' tag aat tC GGC cGG AGA CAG CAC GCA CTT TAC TCT TCC ACC ATT GC 3'). The WD21 primer provides the HindIII restriction site, the aaca sequence promoting the initiation of translation (Joshi, 1987) and the first 8 codons of the transit peptide (including the initiator ATG). The WD22 primer provides the EcoRI restriction site, the first 5 codons of the sequence encoding the mature  $\alpha$  globin chain (an EagI restriction site is created by silent mutation in the fourth codon (CCT modified to CCg)) and the last 7 codons of the sequence of the transit peptide. The PCR amplification and the treatment of the amplified DNA fragments were carried out as described in the third stage of section II.a.

Next, these DNA fragments were doubly digested with HindIII and EcoRI and cloned at the HindIII and EcoRI sites of the plasmid pNEB193 marketed by New England Biolabs. The plasmid pNEB193 was dephosphorylated as described in II.a. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 100 µg/ml ampicillin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC54. The nucleotide

sequence of this chimeric gene resulting from the translational fusion between the sequence encoding the transit peptide and the cDNA encoding the mature  $\alpha$  globin chain was verified by sequencing as described in section I.b.

From the plasmid pBIOC54, the HindIII-EagI fragment, carrying the sequence encoding the transit peptide of the precursor of the small subunit of ribulose 1,5-diphosphate carboxylase and the first 4 codons of the mature  $\alpha$  globin chain was isolated by double digestion, HindIII (total digestion) and EagI (partial digestion). This HindIII-EagI fragment, purified by electrophoresis on a 1.8% agarose gel and by the action of the "Geneclean II" kit (BIO101) was cloned at the HindIII and EagI sites of the dephosphorylated plasmid pBIOC44 as described in section II.a. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 100  $\mu$ g/ml ampicillin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC55. The nucleotide sequence of this chimeric gene resulting from the translational fusion between the sequence encoding the transit peptide and the cDNA encoding the mature  $\alpha$  globin chain was verified by sequencing as described in section I.b.

b. CONSTRUCTION OF THE PLASMID pBIOC57 CONTAINING THE cDNA ENCODING  $\beta$  GLOBIN FOR CHLOROPLAST TARGETING.

To obtain chloroplast targeting of the  $\beta$  globin chain, the sequence encoding the transit peptide of the Pisum sativum L. ribulose 1,5-diphosphate carboxylase small subunit precursor was fused with the first codon of the sequence encoding the mature  $\beta$  globin chain while maintaining the open reading frame. The cleaving sequence between the sequences of the transit peptide and the mature  $\beta$  globin chain is Cys - Val.

The sequence encoding the transit peptide of the ribulose 1,5-diphosphate carboxylase small subunit precursor was amplified by PCR on the plasmid pJIT117 with the aid of the 2 oligodeoxynucleotides, WD 23 (5' gtcattaaataaca ATG GCT TCT ATG ATA TCC TCT TCA GCT GTG 3') and WD24 (5' aatgagct C GGC cGA CTT

CTC CTC AGG AGT CAG GTG CAC GCA CTT TAC TCT TCC ACC 3'). The WD23 primer provides the PacI restriction site, the aaca sequence promoting the initiation of translation (Joshi, 1987) and preceding the first 10 codons of the transit peptide (including the initiator ATG). The WD24 primer provides the SacI restriction site, the first 10 codons of the sequence encoding the mature  $\beta$  globin chain (an EagI restriction site is created by silent mutation in the ninth codon (TCT modified to TCg)) and the last 6 codons of the sequence of the transit peptide. The PCR amplification and the treatment of the amplified DNA fragments were carried out as described in the third stage of section II.a. Next, these DNA fragments were doubly digested with PacI and SacI, purified by electrophoresis on a 1.8% agarose gel and by the action of the "Geneclean II" kit (BIO101) and cloned at the PacI and SacI sites of the plasmid pNEB193 marketed by New England Biolabs. The plasmid pNEB193 was dephosphorylated as described in II.a. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 100  $\mu$ g/ml ampicillin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC56. The nucleotide sequence of this chimeric gene resulting from the translational fusion between the sequence encoding the transit peptide and the cDNA encoding the mature  $\beta$  globin chain was verified by sequencing as described in section I.b.

From the plasmid pBIOC56, the PacI-EagI fragment, carrying the sequence of the transit peptide of the ribulose 1,5-diphosphate carboxylase small subunit precursor and the first 9 codons of the sequence encoding the mature  $\beta$  globin chain, was isolated by double digestion, PacI (total digestion) and EagI (partial digestion). This PacI-EagI fragment, purified by electrophoresis on a 1.8% agarose gel and by the action of the "Geneclean II" kit (BIO101), was cloned at the PacI and EagI sites of the dephosphorylated plasmid pBIOC45 as described in section II.a. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 100  $\mu$ g/ml ampicillin, was extracted according to the alkaline lysis method (Stephen et al., 1990), and analyzed by enzymatic digestion. The resulting clone was called pBIOC57. The nucleotide sequence of this

chimeric gene resulting from the translational fusion between the sequence encoding the transit peptide and the cDNA encoding the mature  $\beta$  globin chain was verified by sequencing as described in section I.b.

c. CONSTRUCTION OF THE CO-EXPRESSION BINARY PLASMID pBIOC59 CONTAINING THE cDNAs ENCODING THE  $\alpha$  AND  $\beta$  GLOBINS, FOR CHLOROPLAST TARGETING.

The HindIII-EcoRI fragment carrying the cDNA encoding the chloroplast targeting  $\alpha$  globin chain was isolated from pBIOC55 described in section IV.a., and ligated to the plasmid DNA of pBIOC43 doubly digested with HindIII and EcoRI and dephosphorylated as described in II.a. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 10  $\mu$ g/ml tetracycline, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC58.

The PacI-MluI fragment carrying the cDNA encoding the chloroplast targeting  $\beta$  globin chain was isolated from pBIOC57 described in section IV.b., and ligated to the plasmid DNA of pBIOC58 doubly digested with PacI and MluI, and dephosphorylated as described in II.a. The ligation and the transformation were carried out as described in section II.c.3 suing the E. coli Sure tet<sup>r</sup> strain. The plasmid DNA of the clones obtained, selected on 10  $\mu$ g/ml tetracycline, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC59.

The nucleotide sequence of the cDNAs encoding the recombinant  $\alpha$  and  $\beta$  globin chains allowing chloroplast targeting was verified by sequencing as described in section I.b. The plasmid DNA of the binary vector pBIOC59 was introduced by direct transformation into the Agrobacterium tumefaciens LBA4404 strain according to the method of Holsters et al. (1978). The validity of the clone selected was verified by enzymatic digestion of the plasmid DNA introduced.

**V. CONSTRUCTION OF THE CHIMERIC GENES ENCODING THE  $\alpha$  AND  $\beta$  GLOBIN CHAINS ALLOWING EXPRESSION OF THE RECOMBINANT HUMAN HEMOGLOBIN FOR SECRETION IN TOBACCO LEAVES.**

To obtain secretion, the sequence encoding the signal peptide (SP) of sporamine A of the tuberized roots of sweet potato (Murakami et al., 1986; Matsuoka and Nakamura, 1991) (ATG AAA GCC TTC ACA CTC GCT CTC TTC TTA GCT CTT TCC CTC TAT CTC CTG CCC AAT CCA GCC CAT TCC), is fused with the first codon of the sequence encoding, on the one hand, the mature  $\alpha$  globin chain (deletion of the initiator ATG) and, on the other hand, the mature  $\beta$  globin chain (deletion of the initiator ATG) while maintaining the open reading frames. This signal peptide of 23 amino acids was isolated from the plasmid pMAT103 (Matsuoka and Nakamura, 1991) and used during the carrying out of the constructions.

a. CONSTRUCTION OF THE PLASMID pBIOC60 CONTAINING THE cDNA ENCODING  $\alpha$  GLOBIN FOR SECRETION.

To obtain the secretion of the  $\alpha$  globin chain, the sequence encoding the signal peptide of sweet potato sporamine A was fused with the first codon of the mature  $\alpha$  globin chain while maintaining the open reading frame. The cleaving sequence between the sequences of the signal peptide and the mature  $\alpha$  globin chain is Ser-Val.

The sequence encoding the signal peptide (SP) of the sporamine A of the tuberized roots of sweet potato was amplified by PCR on the plasmid pMAT103 with the aid of 2 oligodeoxynucleotides, WD 25 (5' cgcaagcttaaca ATG AAA GCC TTC ACA CTC GC 3') and WD36 (5' tagaattC GGC cGG AGA CAG CAC GGA ATG GGC TGG ATT GGG CAG G 3'). The WD25 primer provides the HindIII restriction site, the aaca sequence promoting the initiation of translation (Joshi, 1987) and the first 6 codons of the signal peptide (including the initiator ATG). The WD26 primer provides the EcoRI restriction site, the first 5 codons of the sequence encoding the mature  $\alpha$  globin chain (an EagI restriction site is created by silent mutation in the fourth codon

(CCT modified to CCg)) and the last 7 codons of the sequence of the signal peptide. The PCR amplification and the treatment of the amplified DNA fragments were carried out as described in the third stage of section II.a. Next, these DNA fragments were doubly digested with HindIII and EagI, purified by electrophoresis on a 1.8% agarose gel and by the action of the "Geneclean II" kit (BIO101) and cloned at the HindIII and EagI sites of the dephosphorylated plasmid pBIOC44 described in section II.a. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 100 µg/ml ampicillin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC60. The nucleotide sequence of this chimeric gene resulting from the translational fusion between the sequence encoding the signal peptide and the cDNA encoding the mature  $\alpha$  globin chain was verified by sequencing as described in section I.b.

b. CONSTRUCTION OF THE PLASMID pBIOC61 CONTAINING THE cDNA ENCODING THE  $\beta$  GLOBIN FOR SECRETION.

To obtain the secretion of the  $\beta$  globin chain, the sequence encoding the signal peptide of the sweet potato sporamine A was fused with the first codon of the mature  $\beta$  globin chain while maintaining the open reading frame. The cleaving sequence between the sequences of the signal peptide and the mature  $\beta$  globin chain is Ser-Val.

The sequence encoding the signal peptide (SP) of the sporamine A of the tuberized roots of sweet potato was amplified by PCR on the plasmid pMAT103 with the aid of the 2 oligodeoxynucleotides, WD27 (5' gtcattaaataaca ATG AAA GCC TTC ACA CTC GC 3') and WD28 (5' aatgagct C GGC cGA CTT CTC CTC AGG AGT CAG GTG CAC GGA ATG GGC TGG ATT GGG CAG G 3'). The WD27 primer provides the PacI restriction site, the aaca sequence promoting the initiation of translation (Joshi, 1987) and the first 6 codons of the signal peptide (including the initiator ATG). The WD28 primer provides the SacI restriction site, the first 10 codons of the sequence encoding the mature  $\beta$  globin chain (an EagI site is created by silent

mutation in the ninth codon (TCT modified to TCg)) and the last 7 codons of the sequence of the signal peptide. The PCR amplification and the treatment of the amplified DNA fragments were carried out as described in the third stage of section II.a. Next, these DNA fragments were doubly digested with PacI and EagI, purified by electrophoresis on a 1.8% agarose gel and by the action of the "Geneclean II" kit (BIO101) and cloned at the PacI and EagI sites of the dephosphorylated plasmid pBIOC45 described in section II.b. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 100 µg/ml ampicillin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC61. The nucleotide sequence of this chimeric gene resulting from the translational fusion between the sequence encoding the signal peptide and the cDNA encoding the mature  $\beta$  globin chain was verified by sequencing as described in section I.b.

c. CONSTRUCTION OF THE CO-EXPRESSION BINARY PLASMID pBIOC63  
CONTAINING THE cDNAs ENCODING THE  $\alpha$  AND  $\beta$  GLOBINS FOR  
SECRETION.

The HindIII-EcoRI fragment carrying the cDNA encoding the  $\alpha$  globin chain for secretion was isolated from pBIOC60 described in section V.a., and ligated to the plasmid DNA of pBIOC43 doubly digested with HindIII and EcoRI, and dephosphorylated as described in II.a. The ligation and the transformation were carried out as described in section II.c.3 using the E.coli Sure tet. strain. The plasmid DNA of the clones obtained, selected on 10 µg/ml tetracycline, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC62.

The PacI-MluI fragment carrying the cDNA encoding the  $\beta$  globin chain for secretion was isolated from pBIOC61 described in section V.b., and ligated to the plasmid DNA of pBIOC62 doubly digested with PacI and MluI, and dephosphorylated as described in II.a. The ligation and the transformation were carried out as described

in section I.b. The plasmid DNA of the clones obtained, selected on 10 µg/ml tetracyclin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC63.

The nucleotide sequence of the cDNAs encoding the  $\alpha$  and  $\beta$  globin chains for secretion was verified by sequencing as described in section I.b. The plasmid DNA of the binary vector pBIOC63 was introduced by direct transformation into the *Agrobacterium tumefaciens* LBA4404 strain according to the method of Holsters et al. (1978). The validity of the clone selected was verified by enzymatic digestion of the plasmid DNA introduced.

## **VI. CONSTRUCTION OF THE CHIMERIC GENES ENCODING THE $\alpha$ AND $\beta$ GLOBIN CHAINS ALLOWING EXPRESSION OF RECOMBINANT HUMAN HEMOGLOBIN IN THE ENDOPLASMIC RETICULUM OF TOBACCO LEAVES.**

The sequence encoding the KDEL signal (Lys-Asp-Glu-Leu), placed at the C-terminal end of the  $\alpha$  and  $\beta$  globin chains upstream of the stop codon combined with the presence of the sequence encoding the N-terminal signal peptide (SP) of sporamine A of the tuberized roots of sweet potato allows targeting in the endoplasmic reticulum.

### a. CONSTRUCTION OF THE PLASMID pBIOC65 CONTAINING THE cDNA ENCODING THE $\alpha$ GLOBIN ALLOWING RETENTION IN THE ENDOPLASMIC RETICULUM.

To obtain retention in the endoplasmic reticulum, the sequence encoding the KDEL signal (5' aaa gat gag cta 3') was placed before the first stop codon (TGA) of the mature  $\alpha$  globin chain while maintaining the open reading frame.

The plasmid containing the cDNA encoding the  $\alpha$  globin chain which contains the sequence encoding the KDEL signal placed before its first stop codon was obtained by following the same steps as for the manufacture of the plasmid pBIOC44

described in II.a. except that the WD29 primer (5' gcgaattc TCA tag ctc atc ttt ACG GTA TTT GGA GGT CAG CAC 3') replaces the WD14 primer. The WD29 primer provides the EcoRI restriction site and the KDEL sequence situated respectively after and before the stop codon.

The resulting plasmid obtained was called pBIOC64. The nucleotide sequence of the chimeric gene between the cDNA encoding the  $\alpha$  globin chain and the sequence encoding of the  $\alpha$  KDEL signal was verified by sequencing as described in section I.b.

Next, the plasmid pBIOC64 was modified as described in V.a. by translational fusion with the signal peptide of sporamine A of the tuberized roots of sweet potato to give the plasmid pBIOC65 allowing targeting in the endoplasmic reticulum. The nucleotide sequence of the chimeric gene between the sequence encoding the signal peptide, the cDNA encoding the mature  $\alpha$  globin chain and the sequence encoding the KDEL signal was verified by sequencing as described in section I.b. The cleaving sequence between the sequences of the signal peptide and the mature  $\alpha$  globin chain is Ser - Val.

b. CONSTRUCTION OF THE PLASMID pBIOC67 CONTAINING THE cDNA ENCODING THE  $\beta$  GLOBIN ALLOWING RETENTION IN THE ENDOPLASMIC RETICULUM.

To obtain retention in the endoplasmic reticulum, the sequence encoding the KDEL signal (5' aaa gat gag cta 3') was placed before the first stop codon (TAA) of the mature  $\beta$  globin chain while maintaining the open reading frame.

The plasmid containing the cDNA encoding the  $\beta$  globin chain which contains the sequence encoding the KDEL signal before its first stop codon was obtained by following the same steps as for the manufacture of the plasmid pBIOC45 described in II.b., except that the WD30 primer (5' aatgagctcgtaacgcgt TTA tag ctc atc ttt GTG ATA CTT GTG GGC CAG GGC 3') replaces the WD16 primer. The WD30 primer provides the MluI, HpaI and SacI restriction sites and the KDEL sequence placed respectively after and before the stop codon.

The resulting plasmid obtained was called pBIOC66. The nucleotide sequence of the chimeric gene between the cDNA encoding the  $\beta$  globin chain and the sequence encoding the KDEL signal was verified by sequencing as described in section I.b.

Next, the plasmid pBIOC66 was modified as described in V.b. by translational fusion with the signal peptide of the sporamine A of the tuberized roots of sweet potato to give the plasmid pBIOC67 allowing targeting in the endoplasmic reticulum. The nucleotide sequence of the chimeric gene between the sequence encoding the signal peptide, the cDNA encoding the mature  $\beta$  globin chain and the sequence encoding the KDEL signal was verified by sequencing as described in section I.b. The cleaving sequence between the sequences of the signal peptide and the mature  $\beta$  globin chain is Ser - Val.

c. CONSTRUCTION OF THE CO-EXPRESSION BINARY PLASMID pBIOC69  
CONTAINING THE cDNAs ENCODING THE  $\alpha$  AND  $\beta$  GLOBINS ALLOWING  
RETENTION IN THE ENDOPLASMIC RETICULUM.

The HindIII-EcoRI fragment carrying the cDNA encoding the  $\alpha$  globin chain allowing retention in the endoplasmic reticulum was isolated from pBIOC65 described in section VI.a., and ligated to the plasmid DNA of pBIOC43 doubly digested with HindIII and EcoRI, and dephosphorylated as described in II.a. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 10  $\mu$ g/ml tetracyclin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC68.

The PacI-MluI fragment carrying the cDNA encoding the  $\beta$  globin chain allowing retention in the endoplasmic reticulum was isolated from pBIOC67 described in section VI.b., and ligated to the plasmid DNA of pBIOC68 doubly digested with PacI and MluI, and dephosphorylated as described in II.a. The ligation and the transformation were carried out as described in section II.c.3 using the E. coli Sure tet<sup>r</sup> strain. The plasmid DNA of the clones obtained, selected on 10  $\mu$ g/ml tetracyclin, was

extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC69.

The nucleotide sequence of the cDNAs encoding the  $\alpha$  and  $\beta$  globin chains allowing their retention in the endoplasmic reticulum was verified by sequencing as described in section I.b. The plasmid DNA of the binary vector pBIOC69 was introduced by direct transformation into the Agrobacterium tumefaciens LBA4404 strain according to the method of Holsters et al. (1978). The validity of the clone selected was verified by enzymatic digestion of the plasmid DNA introduced.

## **VII. CONSTRUCTION OF THE CHIMERIC GENES ENCODING THE $\alpha$ AND $\beta$ GLOBIN CHAINS ALLOWING EXPRESSION OF THE RECOMBINANT HUMAN HEMOGLOBIN IN THE VACUOLES OF TOBACCO LEAVES.**

To allow vacuolar targeting, the sequence encoding the prepropeptide (PPS) of sporamine A of the tuberized roots of sweet potato (Murakami et al., 1986; Matsuoka and Nakamura, 1991), which corresponds to the signal peptide followed by the N-terminal sequence for vacuolar targeting (ATG AAA GCC TTC ACA CTC GCT CTC TTC TTA GCT CTT TCC CTC TAT CTC CTG CCC AAT CCA GCC CAT TCC AGG TTC AAT CCC ATC CGC CTC CCC ACC ACA CAC GAA CCC GCC), is fused with the first codon of the sequence encoding, on the one hand, the mature  $\alpha$  globin chain (deletion of the initiator ATG) and, on the other hand, the mature  $\beta$  globin chain (deletion of the initiator ATG) while maintaining the open reading frames. This prepropeptide of 37 amino acids was isolated from the plasmid pMAT103 (Matsuoka and Nakamura, 1991) and used during the carrying out of the constructions.

To obtain vacuolar targeting of the  $\alpha$  globin chain, the sequence encoding the prepropeptide of sweet potato sporamine A was fused with the first codon of the mature  $\alpha$  globin chain while maintaining the open reading frame. The cleaving sequence between the sequences of the signal peptide and the mature  $\alpha$  globin chain is Ala-Val.

a. CONSTRUCTION OF THE PLASMID pBIOC70 CONTAINING THE cDNA ENCODING THE  $\alpha$  GLOBIN ALLOWING VACUOLAR TARGETING.

The sequence encoding the N-terminal prepropeptide (PPS) of the sporamine A of the tuberized roots of sweet potato was amplified by PCR on the plasmid pMAT103 with the aid of the 2 oligodeoxynucleotides, WD25 (5' cgcaagcttaaca ATG AAA GCC TTC ACA CTC GC 3') described in V.a. and WD31 (5' tagaattC GGC cGG AGA CAG CAC GGC GGG TTC GTG TGT GGT TG 3'). The WD31 primer provides the EcoRI restriction site, the first 5 codons of the sequence encoding the mature  $\alpha$  globin chain (an EagI site is created by silent mutation in the fourth codon (CCT modified to CCg)) and the last 6 codons of the sequence of the N-terminal prepropeptide. The PCR amplification and the treatment of the amplified DNA fragments were carried out as described in the third stage of section II.a. Next, these DNA fragments were doubly digested with HindIII and EagI, purified by electrophoresis on a 1.8% agarose gel and by the action of the "Geneclean II" kit (BIO101) and cloned at the HindIII and EagI sites of the dephosphorylated plasmid pBIOC44 described in II.a. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 100  $\mu$ g/ml ampicillin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC70. The nucleotide sequence of the chimeric gene between the sequence encoding the prepropeptide and the cDNA encoding the mature  $\alpha$  globin chain was verified by sequencing as described in section I.b.

b. CONSTRUCTION OF THE PLASMID pBIOC71 CONTAINING THE cDNA ENCODING  $\beta$  GLOBIN ALLOWING VACUOLAR TARGETING.

To obtain vacuolar targeting of the  $\beta$  globin chain, the sequence encoding the prepropeptide of sporamine A of sweet potato was fused with the first codon of the mature  $\beta$  globin chain while maintaining the open reading frame. The cleaving

sequence between the sequences of the signal peptide and the mature  $\beta$  globin chain is Ala-Val.

The sequence encoding the N-terminal prepropeptide (PPS) of sporamine A of the tuberized roots of sweet potato was amplified by PCR on the plasmid pMAT103 with the aid of the 2 oligodeoxynucleotides, WD27 (5' gtcattaattaaca ATG AAA GCC TTC ACA CTC GC 3') described in V.b. and WD32 (5' aatgagct C GGC cGA CTT CTC CTC AGG AGT CAG GTG CAC GGC GGG TTC GTG TGT GGT TG 3'). The WD32 primer provides the SacI restriction site, the first 10 codons of the sequence encoding the mature  $\beta$  globin chain (an EagI restriction site is created by silent mutation in the ninth codon (TCT modified to TCg)) and the last 6 codons of the sequence of the N-terminal prepropeptide. The PCR amplification and the treatment of the amplified DNA fragments were carried out as described in the third stage of section II.a. Next, these DNA fragments were doubly digested with PacI and EagI, purified by electrophoresis on a 1.8% agarose gel and by the action of the "Geneclean II" kit (BIO101) and cloned at the PacI and EagI sites of the dephosphorylated plasmid pBIOC45 described in section II.b. The ligation and the transformation were carried out as described in section I.b. The plasmid DNA of the clones obtained, selected on 100  $\mu$ g/ml ampicillin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC71. The nucleotide sequence of the chimeric gene between the sequence encoding the prepropeptide and the cDNA encoding the mature  $\beta$  globin chain was verified by sequencing as described in section I.b.

c. CONSTRUCTION OF THE CO-EXPRESSION BINARY PLASMID pBIOC73  
CONTAINING THE cDNAs ENCODING THE  $\alpha$  AND  $\beta$  GLOBINS ALLOWING  
VACUOLAR TARGETING.

The HindIII-EcoRI fragment carrying the cDNA encoding the vacuolar targeting  $\alpha$  globin chain was isolated from pBIOC70 described in section VII.a., and ligated to the plasmid DNA of pBIOC43 doubly digested with HindIII and EcoRI, and dephosphorylated as described in II.a. The ligation and the transformation were carried

out as described in section I.b. The plasmid DNA of the clones obtained, selected on 10 µg/ml tetracyclin, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC72.

The PacI-MluI fragment carrying the cDNA encoding the vacuolar targeting β globin chain was isolated from pBIOC71 described in section VII.b., and ligated to the plasmid DNA of pBIOC72 doubly digested with PacI and MluI, and dephosphorylated as described in II.a. The ligation and the transformation were carried out as described in section II.c.3. using the *E. coli* Sure tet<sup>r</sup> strain. The plasmid DNA of the clones obtained, selected on 10 µg/ml tetracycline, was extracted according to the alkaline lysis method (Stephen et al., 1990) and analyzed by enzymatic digestion. The resulting clone was called pBIOC73.

The nucleotide sequence of the cDNAs encoding the vacuolar targeting α and β globin chains was verified by sequencing as described in section I.b. The plasmid DNA of the binary vector pBIOC73 was introduced by direct transformation into the *Agrobacterium tumefaciens* LBA4404 strain according to the method of Holsters et al. (1978). The validity of the clone selected was verified by enzymatic digestion of the plasmid DNA introduced.

## VIII. PRODUCTION OF TRANSGENIC TOBACCO PLANTS

The tobacco plants used for the transformation experiments (*Nicotiana tabacum* var. PBD6) are cultured in vitro on Murashige and Skoog basic medium (1962) supplemented with Gamborg et al. vitamins (1968, Sigma reference M0404), sucrose at 20 g/l and agar (Merck) at 8 g/l. The pH of the medium is adjusted to 5.8 with a solution of potassium hydroxide before autoclaving at 120°C for 20 min. The tobacco plantlets are transplanted by taking internode cuttings every 30 days on this MS20 propagation medium.

All the in vitro cultures are carried out in an air-conditioned chamber under the conditions defined below:

- Light intensity of 30 µE.m<sup>-2</sup>.S<sup>-1</sup>; photoperiod of 16 h;
- Thermoperiod of 26°C by day, 24°C by night.

The transformation technique used is derived from that of Horsch et al. (1985).

A preculture of *Agrobacterium tumefaciens* LBA4404 strain containing the plasmids pBIOC46 or pBIOC47 or pBIOC49 or pBIOC53 or pBIOC59 is carried out for 48 h at 28°C, with stirring, in LB medium supplemented with appropriate antibiotics (rifampicin and tetracycline). The preculture is then diluted 50-fold in the same medium and cultured under the same conditions. After one night, the culture is centrifuged (10 min, 3000 g), the bacteria are taken up in an equivalent volume of liquid MS30 medium (30 g/l sucrose) and this suspension is diluted 10-fold.

Explants of about 1 cm<sup>2</sup> are cut from the leaves of the plantlets described above. They are then brought into contact with the bacterial suspension for 1 h, and then dried rapidly on filter paper and placed on a coculture medium (solid MS30).

After 2 days, the explants are transferred to Petri dishes on MS30 regeneration medium containing a selective agent, kanamycin (200 mg/l), a bacteriostatic, augmentin (400 mg/l) and the hormones necessary for the induction of buds (BAP, 1 mg/l and NAA, 0.1 mg/l). A transplantation of the explants is carried out on the same medium after 2 weeks of culture. After a further 2 weeks, the buds are transplanted into Petri dishes on the development medium composed of the MS20 medium supplemented with kanamycin and augmentin. After 15 days, the buds are transplanted into pots on the same medium whose kanamycin concentration has been decreased by one half. The rooting takes about 20 days, at the end of which the plantlets can be cloned using internode cuttings in vitro or taken out to the greenhouse.

#### **IX: PARTIAL EXTRACTION AND PARTIAL PURIFICATION OF RECOMBINANT PROTEINS FROM TOBACCO LEAVES**

Fifty grams of transformed tobacco leaves (fresh weight) are ground in liquid nitrogen and then left stirring for 15 min at 4°C in 300 ml of 50 mM tris-HCl buffer pH 8 supplemented with 1 mM EDTA, 1 mM β-mercaptoethanol and polyvinylpyrrolidone (PVP, 10 g/300 ml). The ground product is filtered on miracloth and then centrifuged for 20 min at 4°C at 10000 g. The supernatant is again filtered on

miraclot. The proteins are then precipitated for 12 h at 4°C with a solution of ammonium sulfate at saturation. After centrifuging for 20 min at 10000 g, the pellet is taken up in 50 mM tris-HCl buffer pH 8 supplemented with 1 mM DTT and 1 mM EDTA and dialyzed twice 12 hours against this same buffer. After dialysis, the retentate is centrifuged and then filtered on miraclot. An assay of proteins is also carried out according to the Bradford technique (1976).

First purification step: Equilibration in 10 mM phosphate buffer pH 6.7-1 mM EDTA by passing over a Sephadex G25 resin and then loading onto an ion-exchange resin (CM cellulose) equilibrated in 10 mM phosphate buffer pH 6.7, 1 mM EDTA. Washing with 4 volumes of this same buffer and then eluting with a linear gradient from 10 mM Na<sub>2</sub>HPO<sub>4</sub> pH 6.7, 1 mM EDTA to 100 mM Na<sub>2</sub>HPO<sub>4</sub> pH 6.7, 1 mM EDTA.

Second purification step: Equilibration in 10 mM Tris-HCl buffer pH 8.4-1 mM EDTA by passing over a Sephadex G25 resin and then loading onto an ion-exchange resin DEAE-Sephacel equilibrated in 10 mM Tris-HCl buffer pH 8.4, 1 mM EDTA. Washing with 4 volumes of this same buffer and then eluting with a 20 mM KH<sub>2</sub>PO<sub>4</sub> buffer pH 7.4. The pH and ionic strength conditions can be modified according to the nature of the hemoglobin variant.

#### Detection of hemoglobin

Hemoglobin (Hb) is detected by virtue of its chromophore, heme, which gives it its characteristic color. At low concentration and in the presence of another chromophore or molecule which scatters light, the signal due to Hb may be masked. This problem can be overcome using a dynamic technique which makes it possible to detect the presence of Hb in a complex system.

This method is based on differential spectra corresponding to a transition between two forms of Hb and on the photodissociation properties of ligands such as O<sub>2</sub> and CO (Gibson, 1956; Mardenet et al., 1994). The probability of dissociation being higher for CO, this ligand is therefore preferably used. The preparation of the samples is carried out under anaerobic conditions.

The experimental equipment is composed of two sources of light: the first is a pulsed source (laser) which dissociates the ligands, and the second is a continuous lamp which makes it possible to observe the recombination of the ligands by virtue of a change in the intensity of light transmitted (Figures 4, 5 and 6). The photodissociation is efficient in the entire visible spectral domain; our system consists of a YAG laser whose pulses have a duration of 10 ns at 532 nm. The detection is more sensitive in the Soret band (416 nm); we chose 436 nm close to the maximum absorption of the deoxy form. The changes in transmitted intensity occur first of all in a time of the order of the nanosecond (geminate phase) and then continue in a few milliseconds (bimolecular phase). We are particularly interested in this second phase which reflects the allosteric transitions of Hb (Figures 5 and 6). Rapid and reversible kinetic studies make it possible to obtain numerous data and therefore a reliable indication of the state of Hb as regards its normal, physiological function.

The preparation of the samples is carried out as described below. The tobacco leaves (20 g) are ground in liquid nitrogen and then the ground product is mixed with 60 ml of the extraction solution (25 mM Tris-HCl pH 7.5, 10 mM  $\beta$ -mercaptoethanol, 1 mM EDTA). The homogenate is centrifuged at 10000 g at 4°C for 15 minutes. The supernatant containing the soluble proteins is collected. The assay of the proteins is carried out according to the Bradford technique (1976). To 1 ml of plant protein extract (1 mg/ml) are added 32  $\mu$ l and 3.2  $\mu$ l of a concentrated human hemoglobin solution (3.13 mg/ml) in order to obtain solutions containing 100%, 10% and 1% hemoglobin, respectively, relative to the total proteins.

The results obtained are the following:

The kinetics of the samples equilibrated under 0.1 atm CO for three Hb concentrations: 100%, 10% and 1% of the total proteins present in the extract in an amount of 1 mg/ml were measured. The curves are biphasic, similar to those of Hb alone, and exhibit a normal speed (of the order of 1000/s) for the rapid phase (Figure 5). The kinetics are similar for the two concentrations, with the exception of the increase in noise (signal) predictable at low concentration. No signal was observed for the plant extract in the absence of Hb, under the same conditions. We can conclude that the kinetics of recombination of CO with HbA in an extract of tobacco leaves is normal.

## X. EXTRACTION AND PARTIAL PURIFICATION OF RECOMBINANT HEMOGLOBIN FROM TOBACCO SEEDS

In this section, the techniques used for the detection by Western blotting, the extraction and partial purification and the demonstration of the functionality of the recombinant hemoglobin produced in the seeds of transgenic tobacco plants (rHb), are described. The latter are obtained by transformation of the coexpression plasmid pBIOC 59 containing the cDNAs encoding the  $\alpha$  and  $\beta$  globins allowing targeting in the chloroplast.

### a. WESTERN-BLOT DETECTION OF THE RECOMBINANT HEMOGLOBIN ACCUMULATED IN TOBACCO SEEDS

Seventy-five milligrams of tobacco seeds (fresh weight) are ground in liquid nitrogen and then in 600  $\mu$ l of 25 mM ice-cold Tris-HCl buffer pH 8 supplemented with 1 mM EDTA, 1 mM DTT and 1 mM PMSF. The ground product is transferred into an Eppendorf tube and centrifuged at 4°C at 10000 g for 10 min. The supernatant is then concentrated by ultrafiltration with the aid of the micropure .45 and microcon 10 devices (Amicon). The assay of the proteins is carried out according to the Bradford technique (1976) using bovine serum albumin (fraction V) as standard.

The proteins are separated according to their apparent molecular mass by polyacrylamide gel electrophoresis in the presence of SDS according to the Laemmli method (Laemmli, 1970) under reducing conditions. The apparatus used is the Mini-protean II (Bio-Rad). The gel consists of a concentration gel (5% acrylamide, 0.17% bis-acrylamide, 63 mM Tris-HCl pH 6.8, 0.1% SDS) and a separating gel (17% acrylamide) 0.56% bis-acrylamide, 375 mM Tris-HCl pH 8.8, 0.1% SDS). The protein samples are previously diluted with 0.25 volume of loading solution (200 mM Tris-HCl pH 6.8, 400 mM DTT, 40% glycerol, 8% SDS, 0.2% bromophenol blue), then treated at 100°C for 5 min and finally loaded onto the gel. The electrophoresis is carried out in Tris-glycine-SDS buffer (25 mM Tris, 250 mM glycine, 1% SDS) at 25 mA.

After electrophoresis, the proteins are transferred onto a nitrocellulose membrane (BA 85, Schleicher & Schuell) by electrotransfer according to the Towbin et al. technique (1979). The transfer is carried out with the aid of the "mini trans blot module" apparatus (Bio-Rad) at 150 V for 90 min in the presence of the transfer solution (25 mM Tris, 192 mM glycine, 20% methanol). The membrane is rinsed for 5 min at room temperature in 1X PBS (10.4 mM Na<sub>2</sub>HPO<sub>4</sub>, 3.2 mM KH<sub>2</sub>PO<sub>4</sub>, 116 mM NaCl) buffer and then dried.

The presence of the globin chains on the Western-blots is detected using, as primary antibody, a rabbit anti-human hemoglobin immune serum (ref: H-4890, Sigma) and, as secondary antibody, an anti-rabbit IgG monoclonal antibody coupled to alkaline phosphatase (A-8025, Sigma). The revealing is performed using the chromogenic substrate [5-bromo-4-chloro-3-indoyl phosphate/nitro blue tetrazolium (BCIP/NBT)].

The membrane is incubated, with stirring, for 5 min in a TBST buffer solution (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05% Tween 20), and then for at least 30 min in the same solution supplemented with 5% skimmed milk powder (Régilait). The latter solution is replaced, 1/5000 of the volume of anti-hemoglobin immune serum is added and the membrane is incubated for at least 2 hours. It is rinsed 3 times 5 min with TBST solution. The incubation with the secondary antibody is carried out for 1 hour with the anti-rabbit IgG monoclonal antibody diluted 1/10000 in the TBST solution. Next, the membrane is again rinsed 3 times. The alkaline phosphatase activity is revealed by incubating the membrane in the revealing solution (100 mM Tris-HCl pH 9.5, 100 mM NaCl, 5 mM MgCl<sub>2</sub>, 330 µg/ml BCIP, 165 µg/ml NBT). The reaction is stopped by rinsing with water.

Figure 7 represents the Western-blot analysis of the protein composition of the extracts of seeds of tobacco plants transformed or otherwise with the plasmid pBIOC59. The polyclonal antibody recognizes the two normal adult hemoglobin (HbA) globin chains separated during the SDS-PAGE electrophoresis. It is observed that the protein extract of the seeds of the transgenic plant T26-22 differs from that of the control plant in the presence of two polypeptides whose apparent molecular mass is similar to that of the globin chains of HbA and which are recognized by the antibody.

Furthermore, they appear to be represented in an equimolar manner. It can therefore be said that in the seeds, the transgenes encoding the fusion proteins transit peptide- $\alpha$  globins and transit peptide- $\beta$  globin are expressed; the cleaving of the transit peptide would be correctly performed, such that the  $\alpha$  and  $\beta$  globins accumulate. In the seeds of 11 plants, of the 20 tobacco plants transformed independently with the plasmid pBIOC59, the presence of the globins is detected. Expressed as equivalents of HbA, the maximum level of about 0.05% rHb relative to the total soluble proteins extracted is observed for the plant T26-22. It was possible to assess this by comparative Western-blot analysis of HbA concentration ranges in the protein extract of control plant seeds.

b. EXTRACTION AND PARTIAL PURIFICATION OF RECOMBINANT HEMOGLOBIN FROM TOBACCO SEEDS.

The partial purification was carried out using as starting material a mixture of the seeds of transgenic tobacco plants transformed with the plasmid pBIOC59 and expressing the rHb.

Fifteen grams of tobacco seeds (fresh weight) are ground in liquid nitrogen and then in 100 ml of 25 mM ice-cold Tris-HCl buffer pH 8 supplemented with 1 mM EDTA, 1 mM DTT and 1 mM PMSF. The ground product is filtered on miracloth® and then the filtrate is centrifuged at 4°C at 10000 g for 10 min. The supernatant is first saturated with carbon monoxide (CO) and then filtered with a 0.22  $\mu$ m filter and finally concentrated by ultrafiltration with the aid of centriprep 10 devices (Amicon). The concentrate is saturated with CO. Two successive chromatographic steps are carried out (4°C) while monitoring the absorbance values at 280 nm (proteins) and 415 nm (hemoproteins). (i) The concentrate is previously filtered with a 0.22  $\mu$ m filter and then loaded onto a Sephadryl-S100 column (Pharmacia) (2.1 cm x 90 cm) equilibrated with buffer D (9.12 mM Na<sub>2</sub>HPO<sub>4</sub>, 20.88 mM NaH<sub>2</sub>PO<sub>4</sub>, 1 mM DTT, 1 mM EDTA, pH 6.5). The fraction containing the rHb is collected, filtered through a 0.22  $\mu$ m filter and then saturated with CO and finally concentrated as above. Sixty-five percent of the proteins are removed at this stage. (ii) This concentrate is loaded onto the second column, a fast-flow S-sepharose (Pharmacia) (1.1 cm x 10 cm) equilibrated

with buffer D. After washing with 8 volumes of buffer D, an ionic strength gradient is applied (buffer D to buffer D containing 500 mM NaCl). The hemoglobin is eluted at one peak. The fractions containing this peak are combined and the proteins are concentrated as described above. Before and after concentrating, the samples are saturated with CO. This concentrate constitutes the rHb-enriched fraction called FE-rHb. Only 3% of the proteins of the extract now remain in this fraction. To obtain a control for subsequent analyzes, this purification scheme was applied under the same conditions to an extract obtained from 15 g of tobacco seeds not expressing rHb, leading to the production of the fraction called FE-Control.

The presence of the  $\alpha$  and  $\beta$  globins in these fractions was tested for using the Western-blot technique under the conditions described in paragraph X.a. The FE-rHb fraction indeed contains rHb, these two polypeptides being detected (Figure 8).

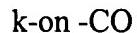
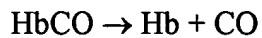
c. DEMONSTRATION OF THE FUNCTIONALITY OF THE RECOMBINANT HEMOGLOBIN BY FLASH PHOTOLYSIS.

The demonstration of the functionality was performed using, as starting material, the rHb-enriched fraction called FE-rHb, using, as control, the equivalent control reaction FE-Control and HbA.

The control experiments where  $1_F$ HbA was added to the plant extract showed biphasic recombination kinetics and variations of the slow fraction depending on the energy of the flash of laser light. These results demonstrate that the function of HbA is not altered by the solvent conditions used.

After photodissociation of the ligands from Hb, bimolecular recombination occurs within a time scale of  $\mu$ s-ms (k-on speed). Although the natural physiological ligand is oxygen, the studies described were performed with carbon monoxide (CO) which gives a photodissociation signal which is much greater than that obtained with  $O_2$  because the yield is higher. Likewise, the difference in the speeds of recombination for the two conformations of Hb (R and T corresponding to tetramers with and without ligand) is also higher. Experimentally, the samples are equilibrated under 0.1 atm CO which gives the best conditions of observation of the two phases. As

the reaction is reversible, the photodissociation ( $\gamma$ ) of the same sample can be repeated in order to accumulate several curves, which greatly improves the signal/noise ratio.



The observation of a variation of the amplitude of slow recombination as a function of the dissociation fraction (by modification of the laser energy) demonstrates the presence of a functional hemoglobin.

The transgenic plants receive genetic information only for the synthesis of globin and not for heme. Consequently, if functional Hb (globins + heme) is expressed in plants, it means that it has captured the heme *in situ*. Other hemoproteins present in plants give an optical signal after flash photolysis. These hemoproteins will not give a signal if the hemin iron is in the ferric form which does not bind the CO and O<sub>2</sub> ligands. CO and O<sub>2</sub> bind reversibly only if the iron atom is in the ferrous form. It is consequently important to demonstrate the existence of kinetic processes for the two phases and the variation in the relative contributions of the two phases due to factors known to influence the function of hemoglobin.

The enriched sample FE-rHb shows a CO photodissociation signal of 48 mOD (optical density) and makes it possible to carry out certain experiments at different levels of dissociation; these experiments are carried out in the absence of sodium dithionite in order to avoid any parasitic contribution due to the presence of hemoproteins. The same experiment, carried out with the FE-Control fraction showed a signal of 1 mOD (Figure 9).

The results recorded at various levels of laser light energy are shown in Figure 10. The curves are similar to that of HbA and shows the existence of a characteristic property of hemoglobin, namely the lower fraction of slow speed when the light intensity is decreased so as to obtain a lower dissociation.

The sample was then equilibrated under a CO atmosphere. As expected, the recombination kinetics are thereafter more rapid. For hemoglobin in solution, the slow fraction is usually lower at high CO concentration since there is less time available to make the R→T transition after dissociation. The FE-rHb sample does not exhibit this effect (Figure 11).

Another method can be used to study the speeds of association and of dissociation of oxygen. The principle of this method is based on the following fact: although CO has an affinity about 200 times higher than that of oxygen, the speed of association of CO with the ligand-containing Hb (R state) is about 10 times lower than for oxygen. A sample equilibrated with an equal mixture of CO and O<sub>2</sub> will be essentially in the HbCO form. It is then possible to photodissociate the CO with a high yield), which allows the study of the recombination of O<sub>2</sub>. A slow terminal phase of the order 1 s due to the replacement of oxygen by CO provides information on the speed of dissociation (k-off). Only the FE-rHb sample reveals a signal for binding of oxygen (Figure 11).

The studies of the FE-rHb fraction by flash photolysis have shown:

- a biphasic recombination of CO with rapid and slow speeds similar to those observed in tetrameric Hb A;
- a decrease in the slow fraction at low laser energy as for Hb A;
- an increase in speed for higher CO concentrations as for normal Hb;
- a reversible binding of oxygen with on and off speeds similar to those of normal Hb A;

It can be concluded that the recombinant hemoglobin produced in tobacco seeds possesses the properties of tetrameric Hb A in all the functional tests carried out.

## **XI. CONSTRUCTION OF CHIMERIC GENES ENCODING THE $\alpha$ AND $\beta$ CHAINS OF HUMAN HEMOGLOBIN AND ALLOWING EXPRESSION IN MAIZE SEEDS.**

**CONSTRUCTION OF THE PLASMIDS CONTAINING ONE OF THE  $\alpha$  OR  $\beta$  CHAINS OF HUMAN HEMOGLOBIN AND ALLOWING CONSTITUTIVE EXPRESSION OR EXPRESSION IN THE ALBUMIN IN MAIZE SEEDS.**

The constitutive or albumin-specific expression, in maize seeds, of the sequences of the  $\alpha$  and  $\beta$  chains of human hemoglobin required the following regulatory sequences:

one of the three promoters allowing a constitutive expression:

- rice actin promoter followed by the rice actin intron (pAR-IAR) contained in the plasmid pAct1-F4 described by McElroy et al. (1991);
- 35S double constitutive promoter (pd35S) of CaMV (cauliflower mosaic virus). It corresponds to a duplication of the sequences activating transcription, situated upstream of the TATA element of the natural 35S promoter (Kay et al., 1987).
- the promoter of the maize  $\gamma$ zein gene (pyzein) contained in the plasmid py63 (Reina et al., 1990). The plasmid py63 results from the cloning of pyzein at the HindIII and XbaI sites of a plasmid pUC18 containing, between its HindIII and EcoRI sites, the expression cassette "p35S-gus-tNOS" of pBI221 marketed by Clontech. It allows expression of the albumin maize seeds. Combined with the rice actin intron, this promoter confers expression is of constitutive type;

one of the two terminators:

- the sequence for termination of transcription, 35S polyA terminator, which corresponds to the noncoding 3' region of the sequence of the circular double-stranded DNA cauliflower mosaic virus producing the 35S transcript (Franck et al., 1980);
- The sequence for termination of transcription, NOS polyA terminator, which corresponds to the noncoding 3' region of the nopaline synthase gene of the Ti plasmid of nopaline-containing strain of Agrobacterium tumefaciens (Depicker et al., 1982).

The type of vector used is derived from pBSIISK+ (Stratagene). Each vector comprises an expression cassette, namely one of the promoters, one of the  $\alpha$  or  $\beta$  chains of human hemoglobin and one of the terminators. Vectors comprising the two

cassettes for expression of each of the  $\alpha$  and  $\beta$  chains of human hemoglobin were also constructed. The clonings were carried out according to the customary methods.

Bibliographic references:

- Benesch & Kwong. Hemoglobin 1994, 18, 185-192.
- Birnboim & Doly. Nucleic Acids Res. 1979, 7, 1513-
- Boutry & Chua. EMBO J. 1985, 4, 2159-2165.
- Bradford. Anal. Biochem. 1976, 72, 248-254.
- Carrer et al. Mol. Gen. Genet. 1993, 241, 49-56.
- Chaumont et al. Plant Mol. Biol. 1994, 24, 631-641.
- Gamborg et al. Exp. Cell Res. 1968, 50, 151-158.
- Guerineau et al. Nucleic Acid Res. 1988, 16, 11380.
- Gibson. J. Physiol. 1956, 134, 123.
- Edelbaum. J. Interferon Res. 1992, 12, 449-453.
- Hanahan. J. Mol. Biol. 1983, 166, 557-
- Hanahan. In "DNA cloning volume I, a practical approach" (Ed: Glover D.M.) IRL Press, 1985, pp 109-135.
- Hiatt & Ma. FEBS Let. 1992, 307, 71-75.
- Hoffman et al. Proc. Natl. Acad. Sci. USA 1990, 87: 8521-8525.
- Horsch et al. Science 1985, 227, 1229-1231.
- International Hemoglobin Information Center (1995) Hemoglobin, 19, 37-124.
- Jessen et al. Meth Enzymol, 1994, 231, 347-364.
- Joshi. Nucleic Acid Res. 1987, 15, 6643-6653.
- Kister et al. J. Biol. Chem. 1987, 262, 12085-12091.
- Krebbers et al. Plant Physiol. 1988, 87, 859-866.
- Marden et al. Meth Enzymol. 1994, 232 71-86.
- Mason et al. Proc. Natl. Acad. Sci. USA 1992, 89, 11745-11749.
- Moloney. Int. Meeting of Production of Recombinant Proteins in Plants, Leicester 1994, page 36-38
- Murashige & Skoog. Physiol. Plantarum 1962, 15, 473-497.

- Nagai & Thogersen. *Meth. Enzymol.* 1987, 153, 461-481.
  - Nagai et al. *Proc. Natl. Acad. Sci. USA* 1985, 82, 7252-7255.
  - Perutz. *Nature* 1970, 228, 726-739.
  - Russel. *Int. Meeting of Production of Recombinant Proteins in Plants*, Leicester 1994, page 43
  - Sambrook et al. *Molecular Cloning: A Laboratory Manual*. Second Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).
  - Sanger et al. *Proc. Natl. Acad. Sci. USA* 1977, 74, 5463-5467
  - Stephen et al. *Nucleic Acid Res.* 1990, 18, 7463-7464.
  - Svab et al. *Proc. Natl. Acad. Sci. USA* 1990, 87, 8526-8530.
  - Swanson et al. *Bio/Technology* 1991, 9, 57-61.
  - Symons et al. *Bio/Technology* 1990, 8, 217-221.
  - Vanderkerckhove et al. *Bio/Technology* 1989, 7, 929-932.
  - Wagenbach et al. *Biotechnology* 1991, 9: 57-61.
  - Wilson et al. *Nucleic Acid Res.* 1978, 5, 563-581.
- Kay R. et al., *Science*, 1987, 236: 1299.
- Franck A. et al., *Cell*, 1980, 21: 285.
- Depicker A. et al., *J. Mol. Appl. Genet.*, 1982, 1: 561.
- Mc Elroy et al., *Mol. Gen. Genet.*, 1991, 231: 150.
- Reina et al., N.A.R., 1990, 18: 6426.
- Dumoulin et al., *Prot. Sci.*, 1996, 5: 114-120.
- Feng et al., *J. Mol. Evol.*, 1985, 21: 112-115.
- Dumoulin et al., *Art. Cells Blood Subs. Immob. Biotech.*, 1994, 22: 733-738.
- Looker et al., *Nature*, 1992, 356: 258-260.
- Laemmli, *Nature*, 1970, 227: 680-685.
- Towbin et al., *Proc. Natl. Acad. Sci. USA*, 1979, 76: 4350-4354.

CLAIMS

1. Method for producing hemin proteins comprising the following steps:
  - i) introducing, into plant cells, one or more nucleic acid molecule(s) each of which comprises at least one sequence encoding a protein component of a hemin protein of animal origin capable of reversibly binding oxygen or a variant or a portion of this protein component, and optionally a sequence encoding a selection agent;
  - ii) selecting the cells containing the nucleic acid encoding the protein component of the hemin protein;
  - iii) optionally propagating the transformed cells, either in culture or by regenerating whole transgenic or chimeric plants;
  - iv) recovering, and optionally purifying, a hemin protein comprising a complex of the protein or proteins encoded by the abovementioned nucleic acid with at least one iron-containing porphyrin nucleus, or a plurality of these complexes.
2. Method according to Claim 1, characterized in that the hemin protein is a single polypeptide chain, for example myoglobin.
3. Method according to Claim 1, characterized in that the hemin protein is a heterodimer, the nucleic acid(s) comprising the sequences encoding each of the various protein units.
4. Method according to Claim 3, characterized in that the hemin protein is human hemoglobin, or a derivative thereof, the nucleic acid(s) comprising sequences encoding  $\alpha$  and  $\beta$  globin, or variants of  $\alpha$  or  $\beta$  globin, the variants differing from the natural sequence in one or more amino acid substitution(s), deletion(s) or insertion(s).
5. Method according to Claim 4, characterized in that the sequences encoding the various protein units,  $\alpha$  and  $\beta$  globin, are contained within the same nucleic acid molecule.
6. Method according to Claim 4, characterized in that the sequences encoding the various protein units,  $\alpha$  and  $\beta$  globin, are contained within separate nucleic acid molecules.

7. Method according to any one of Claims 1 to 6, characterized in that the introduction of the nucleic acid molecule(s) is carried out by transformation of the nuclear genome of the plant cell.
8. Method according to Claim 7, characterized in that the sequence encoding the protein component comprises one or more sequence(s) encoding chloroplast targeting signals, or mitochondrial targeting signals.
9. Method according to Claim 7, characterized in that the sequence encoding the protein component comprises one or more sequence(s) encoding an N-terminal signal peptide and optionally a signal responsible for retaining the protein in the endoplasmic reticulum, or a vacuolar targeting signal.
10. Method according to any one of Claims 1 to 6, characterized in that the introduction of the nucleic acid is performed by a transformation of the mitochondrial or chloroplast genome.
11. Method according to any one of Claims 1 to 9, characterized in that the nucleic acid comprises, in addition to the coding sequence(s), sequences for regulation of transcription which are recognized by plant cells.
12. Method according to any one of Claims 4 to 11, characterized in that the coding sequence(s) encode(s) a hybrid molecule composed of at least the active parts of  $\alpha$ -globin and  $\beta$ -globin.
13. Method according to any one of Claims 1 to 12, characterized in that it comprises, between the propagation step and the recovery step, a step for detecting hemin proteins and in particular hemin proteins whose porphyrin nucleus consists of iron-containing protoporphyrin IX.
14. Method according to Claim 13, characterized in that during the step for propagating plant cells, iron-containing protoporphyrin IX is added to the culture medium used for growing the cells.
15. Hemin protein having the capacity to reversibly bind oxygen, characterized in that it comprises at least one iron-containing porphyrin nucleus, of plant origin, and a protein component comprising at least one polypeptide chain, of animal origin.
16. Protein according to Claim 15, characterized in that the iron-containing porphyrin nucleus is iron-containing protoporphyrin IX, or a protoporphyrin differing

from protoporphyrin IX in the nature of the side chains carried by the  $\beta$  atoms of the pyrrole rings.

17. Protein according to Claim 16, characterized in that the protein component comprises at least one  $\alpha$ - and/or  $\beta$ -globin polypeptide chain, or variants thereof comprising one or more amino acid substitution(s), deletion(s) or insertion(s), the hemin protein being capable of binding oxygen reversibly.

18. Protein according to Claim 17, characterized in that the  $\alpha$ - or  $\beta$ -globin chain, or variants thereof, comprises in addition a chloroplast targeting signal, a mitochondrial targeting signal, or a N-terminal signal peptide optionally in combination with a signal responsible for retaining protein in the endoplasmic reticulum, or a vacuolar targeting signal.

19. Protein according to Claim 17, characterized in that each polypeptide chain lacks NH<sub>2</sub>-terminal methionine.

20. Protein according to any one of Claims 17 to 19, characterized in that the protein component comprises at least four polypeptide chains of  $\alpha$ - and/or  $\beta$ - globin or variants thereof, each polypeptide chain being bound to an iron-containing protoporphyrin nucleus.

21. Protein according to Claim 20, characterized in that it comprises 2  $\alpha$ -globin chains and 2  $\beta$ -globin chains, or variants thereof.

22. Protein according to any one of Claims 15 to 21, characterized in that it binds oxygen with an affinity of between 7 and 40 mm Hg, preferably 15 to 20 mm Hg.

23. Nucleic acid comprising:

- i) one or more sequence(s) encoding a protein component of an animal hemin protein, the said protein having the capacity to reversibly bind oxygen, and
- ii) sequences for regulation of transcription which are recognized by a plant cell, comprising a promoter and sequences for regulation of termination, and
- iii) one or more sequence(s) encoding a targeting signal of plant origin.

24. Nucleic acid according to Claim 23, characterized in that the regulatory sequences comprise one or more promoter(s) of plant origin.

25. Nucleic acid according to Claim 23 or 25, characterized in that the sequences encoding the targeting signal encode a mitochondrial or chloroplast targeting peptide called "transit" peptide.
26. Nucleic acid according to Claim 23 or 24, characterized in that the sequences encoding the targeting signal encode an N-terminal signal peptide of plant origin, optionally in combination with a sequence encoding an endoplasmic retention signal or a vacuolar targeting signal.
27. Nucleic acid according to any one of Claims 23 to 26, characterized in that the coding sequence encodes human  $\alpha$ - or  $\beta$ -globin, or a variant thereof differing from the natural sequence in one or more amino acid substitution(s), deletion(s) or replacement(s), or a portion of human  $\alpha$ - and/or  $\beta$ -globin.
28. Nucleic acid according to any one of Claims 23 to 27, comprising, in addition, one or more intron(s), preferably of plant origin.
29. Nucleic acid according to any one of Claims 23 to 28, characterized in that the sequence encoding the protein component is a cDNA.
30. Vector comprising one or more nucleic acid molecule(s) according to any one of Claims 23 to 29.
31. Plant cells transformed in a stable manner by the nucleic acid according to any one of Claims 23 to 29.
32. Plant cells capable of producing one or more hemin protein(s) according to any one of Claims 15 to 22.
33. Plant cells according to Claim 32, characterized in that they comprise nucleic acid comprising one or more sequence(s) encoding a protein component of the said hemin protein in association with one or more sequence(s) for regulation of transcription recognized by the cell.
34. Plant cells according to any one of Claims 31 to 33, characterized in that they are a culture of plant cells, for example in liquid medium or immobilized cells, or a root culture.
35. Plant cells according to any one of Claims 31 to 33, characterized in that they are cells which form part of a whole transformed plant.

36. Chimeric or transgenic plant capable of producing one or more hemin protein(s), for example hemoglobin or a derivative thereof, characterized in that it comprises cells according to any one of Claims 31 to 33.
37. Seeds of transgenic plant according to Claim 36.
38. Pharmaceutical product comprising one or more hemin protein(s) according to any one of Claims 15 to 22 in association with a physiologically acceptable excipient.
39. Hemin proteins according to any one of Claims 15 to 22 for use as medicament.
40. Hemin protein according to Claim 39, for use in the treatment of conditions requiring an improvement in the transport of oxygen in the blood.
41. Use of a hemin protein according to any one of Claims 15 to 22 for the preparation of a medicament for the treatment of conditions requiring an improvement in the transport of oxygen in the blood.
42. Use of a hemin protein according to any one of Claims 15 to 22 in an industrial, cosmetic product or as chemical reagent.

QdC  
A2

## ABSTRACT

A method for producing haemin proteins by (i) inserting into plant cells one or more nucleic acid molecules that each comprise at least one sequence coding for a protein component of an animal haemin protein capable of reversibly binding oxygen, or for a variant or portion of said protein component, and optionally a sequence coding for a selection agent; (ii) selecting cells containing nucleic acid coding for the protein component of the haemin protein; (iii) optionally propagating the transformed cells either in a culture or by regenerating whole transgenic or chimeric plants; and (iv) recovering and optionally purifying a haemin protein that includes a complex consisting of the protein or proteins coded for by said nucleic acid and at least one iron-containing porphyrinic nucleus, or a plurality of such complexes.